

## EXPLORING THE VARIABLE SKY WITH THE SLOAN DIGITAL SKY SURVEY

BRANIMIR SESAR,<sup>1</sup> ŽELJKO IVEŽIĆ,<sup>1</sup> ROBERT H. LUPTON,<sup>2</sup> MARIO JURIĆ,<sup>3</sup> JAMES E. GUNN,<sup>2</sup> GILLIAN R. KNAPP,<sup>2</sup>  
 NATHAN DE LEE,<sup>4</sup> J. ALLYN SMITH,<sup>5</sup> GAJUS MIKNAITIS,<sup>6</sup> HUAN LIN,<sup>6</sup> DOUGLAS TUCKER,<sup>6</sup> MAMORU DOI,<sup>7</sup>  
 MASAYUKI TANAKA,<sup>8</sup> MASATAKA FUKUGITA,<sup>9</sup> JON HOLTZMAN,<sup>10</sup> STEVE KENT,<sup>6</sup> BRIAN YANNY,<sup>6</sup>  
 DAVID SCHLEGEL,<sup>11</sup> DOUGLAS FINKBEINER,<sup>12</sup> NIKHIL PADMANABHAN,<sup>11</sup> CONSTANCE M. ROCKOSI,<sup>13</sup>  
 NICHOLAS BOND,<sup>2</sup> BRIAN LEE,<sup>11</sup> CHRIS STOUGHTON,<sup>6</sup> SEBASTIAN JESTER,<sup>14</sup> HUGH HARRIS,<sup>15</sup>  
 PAUL HARDING,<sup>16</sup> JON BRINKMANN,<sup>17</sup> DONALD P. SCHNEIDER,<sup>18</sup> DONALD YORK,<sup>19</sup>  
 MICHAEL W. RICHMOND,<sup>20</sup> AND DANIEL VANDEN BERK<sup>18</sup>

*Received 2007 March 23; accepted 2007 July 19*

### ABSTRACT

We quantify the variability of faint unresolved optical sources using a catalog based on multiple SDSS imaging observations. The catalog covers SDSS stripe 82, which lies along the celestial equator in the southern Galactic hemisphere ( $22^{\text{h}}24^{\text{m}} < \alpha_{\text{J2000.0}} < 04^{\text{h}}08^{\text{m}}$ ,  $-1.27^\circ < \delta_{\text{J2000.0}} < +1.27^\circ$ ,  $\sim 290 \text{ deg}^2$ ), and contains 34 million photometric observations in the SDSS *ugriz* system for 748,084 unresolved sources at high Galactic latitudes ( $b < -20^\circ$ ) that were observed at least four times in each of the *ugri* bands (with a median of 10 observations obtained over  $\sim 6$  yr). In each photometric bandpass we compute various low-order light-curve statistics, such as rms scatter,  $\chi^2$  per degree of freedom, skewness, and minimum and maximum magnitude, and use them to select and study variable sources. We find that 2% of unresolved optical sources brighter than  $g = 20.5$  appear variable at the 0.05 mag level (rms) simultaneously in the *g* and *r* bands (at high Galactic latitudes). The majority (2 out of 3) of these variable sources are low-redshift ( $z < 2$ ) quasars, although they represent only 2% of all sources in the adopted flux-limited sample. We find that at least 90% of quasars are variable at the 0.03 mag level (rms) and confirm that variability is as good a method for finding low-redshift quasars as the UV excess color selection (at high Galactic latitudes). We analyze the distribution of light-curve skewness for quasars and find that it is centered on zero. We find that about one-fourth of the variable stars are RR Lyrae stars, and that only 0.5% of stars from the main stellar locus are variable at the 0.05 mag level. The distribution of light-curve skewness in the  $g - r$  versus  $u - g$  color-color diagram on the main stellar locus is found to be bimodal (with one mode consistent with Algol-like behavior). Using over 600 RR Lyrae stars, we demonstrate rich halo substructure out to distances of 100 kpc. We extrapolate these results to the expected performance by the Large Synoptic Survey Telescope and estimate that it will obtain well-sampled, 2% accurate, multicolor light curves for  $\sim 2$  million low-redshift quasars and discover at least 50 million variable stars.

**Key words:** Galaxy: halo — Galaxy: stellar content — quasars: general — stars: Population II — stars: variables: other

**Online material:** color figures

### 1. INTRODUCTION

Variability is an important phenomenon in astrophysical studies of structure and evolution, both stellar and galactic. Some variable stars, such as RR Lyrae stars, are an excellent tool for studying the Galaxy. Being nearly standard candles (thus making distance determination relatively straightforward) and being intrinsically bright, they are a particularly suitable tracer of Galactic structure. In extra-

galactic astronomy, the optical continuum variability of quasars is utilized as an efficient method for their discovery (van den Bergh et al. 1973; Hawkins 1983; Koo et al. 1986; Hawkins & Veron 1995) and is also frequently used to constrain the origin of their emission (Kawaguchi et al. 1998; Trevese et al. 2001; Martini & Schneider 2003).

Despite the importance of variability, the variable optical sky remains largely unexplored and poorly quantified, especially at

<sup>1</sup> Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195-1580, USA.

<sup>2</sup> Princeton University Observatory, Princeton, NJ 08544-1001, USA.

<sup>3</sup> Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA.

<sup>4</sup> Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824-2320, USA.

<sup>5</sup> Department of Physics and Astronomy, Austin Peay State University, Box 4608, Clarksville, TN 37044, USA.

<sup>6</sup> Fermi National Accelerator Laboratory, Box 500, Batavia, IL 60510, USA.

<sup>7</sup> Institute of Astronomy, University of Tokyo, 2-21-1 Osawa, Mitaka, Tokyo 181-0015, Japan.

<sup>8</sup> Department of Astronomy, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

<sup>9</sup> Institute for Cosmic Ray Research, University of Tokyo, Kashiwa, Chiba, Japan.

<sup>10</sup> New Mexico State University, 1320 Frenger Street, Box 30001, Las Cruces, NM 88003, USA.

<sup>11</sup> Lawrence Berkeley National Laboratory, One Cyclotron Road, MS 50R5032, Berkeley, CA 94720, USA.

<sup>12</sup> Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA.

<sup>13</sup> University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95060, USA.

<sup>14</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany.

<sup>15</sup> US Naval Observatory, Flagstaff Station, Box 1149, Flagstaff, AZ 86002, USA.

<sup>16</sup> Department of Astronomy, Case Western Reserve University, Cleveland, OH 44106, USA.

<sup>17</sup> Apache Point Observatory, 2001 Apache Point Road, Box 59, Sunspot, NM 88349-0059, USA.

<sup>18</sup> Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA 16802, USA.

<sup>19</sup> Astronomy and Astrophysics Center, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA.

<sup>20</sup> Department of Physics, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623-5603, USA.

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>DEC 2007</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2007 to 00-00-2007</b>	
4. TITLE AND SUBTITLE <b>Exploring the Variable Sky with the Sloan Digital Sky Survey</b>			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>US Naval Observatory,Flagstaff Station,Box 1149,Flagstaff,AZ,86002</b>			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT <b>see report</b>					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>16</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

the faint end. To what degree different variable populations contribute to the overall variability, how they are distributed in magnitude and color, and what the characteristic timescales and dominant mechanisms of variability are, are just some of the questions that still remain to be answered. To address these questions, several contemporary projects aimed at regular monitoring of the optical sky were started. Some of the more prominent surveys in terms of sky coverage, depth, and cadence are as follows:

1. The Faint Sky Variability Survey (Groot et al. 2003) is a very deep ( $17 < V < 24$ ) *BVI* survey of  $23 \text{ deg}^2$  of sky, containing about 80,000 sources sampled at timescales ranging from minutes to years.
2. The QUEST Survey (Vivas et al. 2004) monitors  $700 \text{ deg}^2$  of sky from  $V = 13.5$  to a limit of  $V = 21$ .
3. ROTSE-I (Akerlof et al. 2000) monitors the entire observable sky twice a night from  $V = 10$  to a limit of  $V = 15.5$ . The Northern Sky Variability Survey (Woźniak et al. 2004) is based on ROTSE-I data.
4. The All Sky Automated Survey (Pojmański 2002) monitors the entire southern and part of the northern sky ( $\delta < 25^\circ$ ) to a limit of  $V = 15$ .
5. OGLE (OGLE II; Udalski et al. 2002) monitors  $\sim 100 \text{ deg}^2$  toward the Galactic bulge from  $I = 11.5$  to a limit of  $I = 20$ . Due to the very high stellar density toward the bulge, OGLE II has detected about 270,000 variable stars (Woźniak et al. 2002; Zeburń et al. 2001).
6. The MACHO Project monitored the brightness of  $\sim 60$  million stars in  $\sim 90 \text{ deg}^2$  of sky toward the Magellanic Clouds and the Galactic bulge for  $\sim 7$  yr to a limit of  $V \sim 24$  (Alcock et al. 2001).

A comprehensive review of past and ongoing variability surveys can be found in Becker et al. (2004).

Recognizing the outstanding importance of variable objects, the last Decadal Survey Report (National Research Council 2001) highly recommended a major new initiative for studying the variable sky: the large survey telescope. The two most ambitious proposals for such a telescope are the Pan-STARRS project (Kaiser et al. 2002) and the Large Synoptic Survey Telescope<sup>21</sup> (LSST; Tyson 2002; Walker 2003). The initial version of Pan-STARRS, with the first 1.8 m telescope (four are planned), has already had its first light, and the 8.4 m LSST will have its first light in 2014 (if approved for construction in 2009).

LSST will offer an unprecedented view of the faint variable sky; according to the current designs it will scan the entire accessible sky every three nights to a limit of  $V \sim 25$  with two observations per night in two different bands (selected from a set of six). One of the LSST science goals<sup>22</sup> will be the exploration of the transient optical sky: the discovery and analysis of rare and exotic objects (e.g., neutron star and black hole binaries), gamma-ray bursts, X-ray flashes, and new classes of transients, such as binary mergers and stellar disruptions by black holes. The observed volume of space, and the requirement to recognize and monitor these events in real time on a “normally” variable sky, will present a challenge to the project.

Since LSST will utilize<sup>23</sup> the Sloan Digital Sky Survey (SDSS; York et al. 2000) photometric system (*ugriz*; Fukugita et al. 1996), multiple photometric observations obtained by the SDSS represent an excellent data set for a pre-LSST study that charac-

terizes the faint variable sky and quantifies the variable population and its distribution in magnitude-color-variability space. Here we present such a study of unresolved sources in a region that has been imaged multiple times by the SDSS.

In § 2 we give a brief overview of the SDSS imaging survey and repeated scans of an  $\sim 290 \text{ deg}^2$  region called “stripe 82.” In § 3 we describe methods used to select candidate variable sources from the SDSS stripe 82 data assembled, averaged, and recalibrated by Ivezić et al. (2007) and present tests that show the robustness of the adopted selection criteria. In the same section we discuss the distribution of selected variable sources in magnitude-color-variability space. The Milky Way halo structure traced by selected candidate RR Lyrae stars is discussed in § 4, and in § 5 we estimate the fraction of variable quasars. Implications for surveys such as the LSST are discussed in § 6, and our main results are summarized in § 7.

## 2. OVERVIEW OF THE SDSS IMAGING AND STRIPE 82 DATA

The quality of the photometry and astrometry, as well as the large area covered by the survey, makes the SDSS stand out among available optical sky surveys (Sesar et al. 2006). The SDSS provides homogeneous and deep ( $r < 22.5$ ) photometry in five band-passes (*u*, *g*, *r*, *i*, and *z*; Gunn et al. 1998, 2006; Hogg et al. 2001; Smith et al. 2002; Tucker et al. 2006) accurate to 0.02 mag (rms scatter) for unresolved sources not limited by photon statistics (Scranton et al. 2002; Ivezić et al. 2003) and with a zero-point uncertainty of 0.02 mag (Ivezić et al. 2004a). The survey sky coverage of  $10,000 \text{ deg}^2$  in the northern Galactic cap and  $300 \text{ deg}^2$  in the southern Galactic cap results in photometric measurements for well over 100 million stars and a similar number of galaxies (Stoughton et al. 2002). The recent Data Release 5 (Adelman-McCarthy et al. 2007)<sup>24</sup> lists photometric data for 215 million unique objects observed in  $8000 \text{ deg}^2$  of sky as part of the “SDSS-I” phase that ran through 2005 June. Astrometric positions are accurate to better than  $0.1''$  per coordinate (rms) for sources with  $r < 20.5$  (Pier et al. 2003), and the morphological information from the images allows reliable star-galaxy separation to  $r \sim 21.5$  (Lupton et al. 2002). In addition, the five-band SDSS photometry can be used for very detailed source classification, e.g., separation of quasars and stars (Richards et al. 2002), spectral classification of stars to within one to two spectral subtypes (Lenz et al. 1998; Finlator et al. 2000; Hawley et al. 2002), and even remarkably efficient color selection of horizontal-branch and RR Lyrae stars (Yanny et al. 2000; Sirko et al. 2004; Ivezić et al. 2005) and low-metallicity G and K giants (Helmi et al. 2003).

The equatorial stripe 82 region ( $22^{\text{h}}24^{\text{m}} < \alpha_{\text{J2000.0}} < 04^{\text{h}}08^{\text{m}}$ ,  $-1.27^\circ < \delta_{\text{J2000.0}} < +1.27^\circ$ ,  $\sim 290 \text{ deg}^2$ ) from the southern Galactic cap ( $-64^\circ < b < -20^\circ$ ) presents a valuable data source for variability studies. The region was repeatedly observed (58 imaging runs from 1998 September to 2004 December, but not all cover the entire region), and it is the largest source of multi-epoch data in the SDSS-I phase. Observations are fairly homogeneously distributed over the 6 yr period, with one to two observations (separated by about a week, on average) obtained every fall. A histogram of the number of observations per star is shown in Figure 1 in Ivezić et al. (2007). Another source for the large number of scans is the SDSS-II Supernova Survey (J. A. Frieman et al. 2007, in preparation). The SDSS-II Supernova Survey scans at least once a week during 4 month long seasons, and two out of three planned seasonal campaigns are already

<sup>21</sup> See <http://www.lsst.org>.

<sup>22</sup> For more details, see [http://www.lsst.org/Science/science\\_goals.shtml](http://www.lsst.org/Science/science_goals.shtml).

<sup>23</sup> LSST will also use the *Y* band at  $\sim 1 \mu\text{m}$ . For more details, see the LSST Science Requirements Document at [http://www.lsst.org/Science/lsst\\_baseline.shtml](http://www.lsst.org/Science/lsst_baseline.shtml).

<sup>24</sup> See <http://www.sdss.org/dr5>.

completed (the SDSS-II Supernova Survey data were not used in this work). By averaging the repeated observations of stripe 82 sources, more accurate photometry than the nominal 0.02 mag single-scan accuracy can be achieved, as demonstrated by Ivezić et al. (2007), who produced a catalog of recalibrated, co-added stripe 82 observations. The catalog lists 58 million photometric observations for 1.4 million unresolved sources that were observed at least four times in each of the *gri* bands (with a median of 10 and a maximum of 28 observations obtained over  $\sim 6$  yr). The random photometric errors for PSF (point-spread function) magnitudes are below 0.01 mag for stars brighter than 19.5, 20.5, 20.5, 20, and 18.5 in *ugriz*, respectively (about twice as accurate for individual SDSS runs), and the spatial variation of photometric zero points is not larger than  $\sim 0.01$  mag (rms). Following Ivezić et al. (2007), we use PSF magnitudes because they go deeper at a given signal-to-noise ratio than aperture magnitudes and have more accurate photometric error estimates than model magnitudes. In addition, various low-order statistics, such as rms scatter ( $\Sigma$ ),  $\chi^2$  per degree of freedom ( $\chi^2$ ), light-curve skewness ( $\gamma$ ), and minimum and maximum PSF magnitude, were computed for each *ugriz* band and each source. We compute  $\chi^2$  per degree of freedom as

$$\chi^2 = \frac{1}{n-1} \sum_{i=1}^n \frac{(x_i - \langle x \rangle)^2}{\xi_i^2} \quad (1)$$

and light-curve skewness  $\gamma$  as<sup>25</sup>

$$\gamma = \frac{n^2}{(n-1)(n-2)} \frac{\mu_3}{\Sigma^3}, \quad (2)$$

$$\mu_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \langle x \rangle)^3, \quad (3)$$

$$\Sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \langle x \rangle)^2}, \quad (4)$$

where  $n$  is the number of detections,  $x_i$  is the magnitude,  $\langle x \rangle$  is the mean magnitude, and  $\xi_i$  is the photometric error.

Separation of quasars and stars, as well as efficient color selection of horizontal-branch and RR Lyrae stars, depends on accurate *u*-band photometry. To ensure this, we select 748,084 unresolved sources from the Ivezić et al. (2007) catalog with at least four detections in the *u* band. Although this cut reduces the initial sample by about a factor of 2, it does not introduce a significant bias in the selection of candidate variable sources, as we show in § 3.

### 3. ANALYSIS OF THE STRIPE 82 CATALOG OF VARIABLE SOURCES

In this section we describe methods for selecting candidate variable sources and present tests that show the robustness of the adopted selection criteria. The distribution of selected variable sources in magnitude-color-variability space is also presented.

#### 3.1. Methods and Selection Criteria

Due to a relatively small number of observations per source and random sampling, we do not perform light-curve fitting but instead use low-order statistics to select candidate variables and study their properties. There are four parameters (median PSF

magnitude, rms scatter  $\Sigma$ ,  $\chi^2$ , and light-curve skewness  $\gamma$ ) measured in five photometric bands (*u*, *g*, *r*, *i*, and *z*), for a total of 20 parameters. In the analysis presented here, we utilize eight of them, as follows:

1. Median PSF magnitudes in the *ugr* bands (corrected for interstellar extinction using the map from Schlegel et al. 1998) because the *g* – *r* versus *u* – *g* color-color diagram has the most classification power (e.g., Smolčić et al. 2004 and references therein).

2.  $\Sigma$  and  $\chi^2$  in the *g* and *r* bands.

3. Light-curve skewness  $\gamma(g)$  (the *g* band combines a high signal-to-noise ratio and large variability amplitude for the majority of variable sources).

The observed rms scatter  $\Sigma$  includes both the intrinsic variability  $\sigma$  and the mean photometric error  $\langle \xi(m) \rangle$  as a function of magnitude. The dependence of  $\Sigma$  on magnitude in the *ugr* bands is shown in Figure 1. For sources brighter than 18, 19.5, 19.5, 19, and 17.5 mag in *ugriz*, respectively, the SDSS delivers 2% photometry with little or no dependence on magnitude. We determine  $\langle \xi(m) \rangle$  by fitting a fourth-degree polynomial to median  $\Sigma$  values in 0.5 mag wide bins (here we assume that the majority of sources are not variable). The theoretically expected  $\langle \xi(m) \rangle$  function (Strateva et al. 2001),

$$\langle \xi(m) \rangle = a + b10^{0.4m} + c10^{0.8m}, \quad (5)$$

provides equally good fits. We define the intrinsic variability  $\sigma$  (hereafter rms scatter  $\sigma$ ) as

$$\sigma = [\Sigma^2 - \langle \xi(m) \rangle^2]^{1/2} \quad (6)$$

for  $\Sigma > \langle \xi(m) \rangle$  and  $\sigma = 0$  otherwise.

As the first variability selection criterion, we adopt  $\sigma(g) \geq 0.05$  and  $\sigma(r) \geq 0.05$  mag [hereafter written as  $\sigma(g, r) \geq 0.05$  mag]. At the bright end, this criterion is equivalent to selecting sources with rms scatter greater than  $2.5\sigma_0$ , where  $\sigma_0 = 0.02$  mag is the measurement noise. Selection cuts are applied simultaneously in the *g* and *r* bands to reduce the number of “false positives” (intrinsically nonvariable sources selected as candidate variable sources due to measurement noise). About 6% of sources pass the  $\sigma$  cut in each band separately, and  $\sim 3\%$  of sources pass the cut in both bands simultaneously. By selecting sources with  $\sigma(g, r) \geq 0.05$  mag, we also select faint sources that have large  $\sigma$  due to large photometric errors at the faint end. To only select faint sources with statistically significant rms scatter, we apply the  $\chi^2$  test as the second selection cut.

In the  $\chi^2$  test, the value of  $\chi^2$  per degree of freedom (calculated with respect to a weighted mean magnitude and using errors computed by the photometric pipeline) determines whether the observed light curve is consistent with the Gaussian distribution of errors. Large  $\chi^2$  values show that the rms scatter is inconsistent with random fluctuations. Ivezić et al. (2003, 2007) used multi-epoch SDSS observations to show that the photometric error distribution in the SDSS roughly follows a Gaussian distribution. A comparison of  $\chi^2$  distributions in the *g* and *r* bands with a reference Gaussian  $\chi^2$  distribution is shown in Figure 2. As is evident,  $\chi^2$  distributions in both bands roughly follow the reference Gaussian  $\chi^2$  distribution for  $\chi^2 < 1$ , demonstrating that median photometric errors are correctly determined. The discrepancy for larger  $\chi^2$  is due to variable sources rather than non-Gaussian error distributions, as we demonstrate below.

<sup>25</sup> We use equations from [http://www.xycoon.com/skewness\\_mall\\_sample\\_test\\_1.htm](http://www.xycoon.com/skewness_mall_sample_test_1.htm).

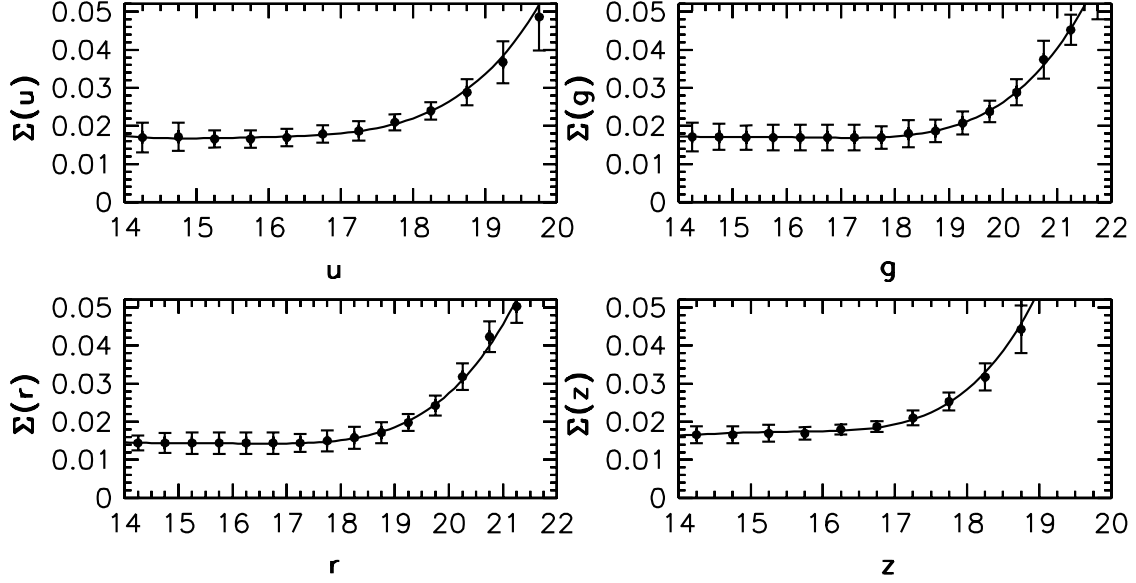


FIG. 1.— Dependence of the median rms scatter  $\Sigma$  in SDSS  $ugrz$  bands on magnitude (symbols). The vertical bars show the rms scatter of  $\Sigma$  in each bin (not the error of the median). The dependence of  $\Sigma$  in the  $i$  band is similar to the  $r$ -band dependence. In each band a fourth-degree polynomial is fitted through the medians (solid line). [See the electronic edition of the *Journal* for a color version of this figure.]

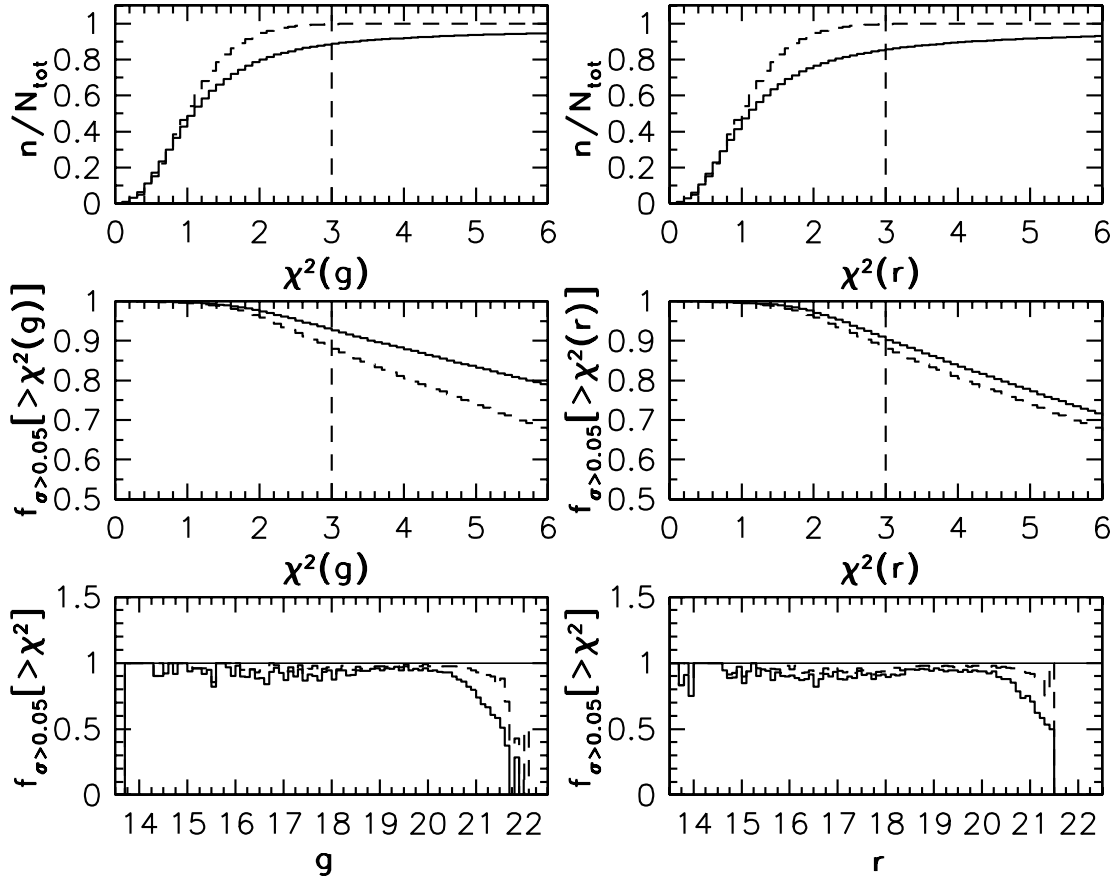


FIG. 2.— *Top panels:* Cumulative distribution of  $\chi^2$   $g$  and  $r$  values for all sources (solid line) and a reference Gaussian  $\chi^2$  distribution with 9 degrees of freedom (dashed line). Dashed lines show adopted selection cuts on  $\chi^2(g)$  and  $\chi^2(r)$  values. *Middle panels:* Fraction of  $\sigma(g, r) \geq 0.05$  mag sources with  $\chi^2$  per degree of freedom greater than  $\chi^2$  (only in the  $g$  or  $r$  band, solid line; in both the  $g$  and  $r$  bands, dashed line). *Bottom panels:* Fraction of  $\sigma(g, r) \geq 0.05$  mag sources with  $\chi^2(m) \geq 2$  (dashed line) or  $\chi^2(m) \geq 3$  (solid line) as a function of magnitude for the  $m = g, r$  bands, respectively. [See the electronic edition of the *Journal* for a color version of this figure.]

The second selection cut,  $\chi^2(g) \geq 3$  and  $\chi^2(r) \geq 3$  [hereafter written as  $\chi^2(g, r) \geq 3$ ], selects  $\sim 90\%$  of  $\sigma(g, r) \geq 0.05$  mag sources, as shown in Figure 2 (*middle panels*). The effectiveness of the  $\chi^2$  test is demonstrated in Figure 2 (*bottom panels*). For magnitudes fainter than  $g = 20.5$ , the fraction of candidate variables decreases as photometric errors increase. The selection is relatively uniform for sources brighter than  $g = 20.5$ , and we adopt this value as the flux limit for the selected variable sample.

There are 662,195 sources brighter than  $g = 20.5$  in the full sample. Using  $\sigma(g, r) \geq 0.05$  mag and  $\chi^2(g, r) \geq 3$  as the selection criteria, we select 13,051 candidate variable sources.<sup>26</sup> Therefore, *at least 2% of unresolved optical sources brighter than  $g = 20.5$  appear variable at the  $\geq 0.05$  mag level (rms) simultaneously in the  $g$  and  $r$  bands* (at high Galactic latitudes).

The required minimum number of  $u$ -band detections (four) imposed on the initial sample rejects 972 sources (or  $\sim 7\%$  of the variable sample). The majority of these sources are redder than  $g - r \sim 1$  and are very faint in the  $u$  band ( $u \sim 22$ ). Our results and conclusions presented below do not change if these sources are included.

We also investigate how the fraction of selected variable sources changes as a function of the minimum required number of observations by requiring a minimum of eight detections in the  $ugr$  bands. The fraction of selected variable sources remains at 2%. We conclude that the fraction of selected variable sources does not depend strongly on the minimum required number of observations.

The fraction of selected variable sources depends on the stellar density because the number of stars increases at lower Galactic latitudes (see Fig. 5 in Ivezić et al. 2007), while the quasar counts remain the same. The detailed makeup of the variable stellar population presumably depends on Galactic coordinates due to a varying stellar population mix.

### 3.2. The Counts of Variable Sources

In this section we estimate the completeness and efficiency of the candidate variable sample and discuss the dependence of counts, rms scatter,  $\sigma(g)/\sigma(r)$  ratio, and light-curve skewness  $\gamma(g)$  on position in the  $g - r$  versus  $u - g$  color-color diagram.

#### 3.2.1. Completeness

The selection completeness, defined as the fraction of true variable sources recovered by the algorithm, depends on the light-curve shape and amplitudes. Due to a fairly large number of observations (median of 10) and small  $\sigma(g, r)$  cutoff compared to typical amplitudes of variable sources (e.g., most RR Lyrae stars and quasars have peak-to-peak amplitudes of  $\sim 1$  mag), we expect the completeness to be fairly high for RR Lyrae stars ( $\geq 95\%$ ; see § 4) and quasars ( $\sim 90\%$ ; see § 5). The completeness for other types of variable sources, such as flares and eclipsing binaries, is hard to estimate but is probably low due to sparse sampling.

#### 3.2.2. Efficiency

The selection efficiency, defined as the fraction of true variable sources in the candidate variable sample, determines the robustness of the selection algorithm. The main diagnostic for the robustness of the adopted selection criteria is the distribution of selected candidates in the SDSS color-magnitude and color-color diagrams. The position of a source in these diagrams is a good proxy for its spectral classification (Lenz et al. 1998; Fan 1999; Finlator et al. 2000; Smolčić et al. 2004).

Figure 3 compares the distribution of candidate variable sources to that of all sources in the  $g - r$  versus  $u - g$  color-color diagram. Were the selection a random process, the selected candidates would have the same distribution as the full sample. The distributions of candidate variables and of the full sample are remarkably different, demonstrating that the candidate variables are *not* randomly selected from the parent sample, and therefore suggesting high selection efficiency.

#### 3.2.3. Dominant Classes of Variable Objects

The three dominant classes of variable objects are quasars, RR Lyrae stars, and stars from the main stellar locus. The most obvious difference between the variable and full sample distributions is a much higher fraction of low-redshift quasars ( $< 2.2$ , recognized by their UV excess,  $u - g < 0.7$ ; see Richards et al. 2002) and RR Lyrae stars ( $u - g \sim 1.15$ ,  $g - r < 0.3$ ; see Ivezić et al. 2005) in the variable sample, vividly shown in Figure 3 (*bottom*).

Another interesting feature visible in Figure 3 (*bottom*) is a gradient in the fraction of variable main stellar locus stars (perpendicular to the main stellar locus). We investigate this gradient by first defining principal colors,

$$P_1 = 0.91u - 0.495g - 0.415r - 1.28 \quad (7)$$

and

$$s = -0.249u + 0.794g - 0.555r + 0.234, \quad (8)$$

where  $P_1$  and  $s$  are the principal axis parallel and perpendicular to the main stellar locus, respectively (Ivezić et al. 2004a). The  $s$  color is a measure of metallicity (Lenz et al. 1998), and  $s > 0.05$  stars are expected to be metal-poor (Helmi et al. 2003). Sources with  $r < 19$  and  $0 < P_1 < 0.9$  are selected and binned in four  $s$  bins. For each bin we calculate the fraction of sources with  $\sigma(g) \geq 0.05$  mag, the fraction of variable sources [selected with  $\sigma(g, r) \geq 0.05$  mag and  $\chi^2(g, r) \geq 3$ ], the median  $\sigma(g)$ , and the total number of sources in the bin (see Table 1). A greater fraction of variable sources in the last bin ( $s > 0.06$ ) indicates that, on average, metal-poor main stellar locus stars are more variable than metal-rich stars. We speculate that this increased variability could be because this sample of metal-poor stars is expected to have a high fraction of giants.

In order to quantify the differences between the full and variable samples, we follow Sesar et al. (2006) and divide the  $g - r$  versus  $u - g$  color-color diagram into six characteristic regions, each dominated by a particular type of source, as shown in Figure 4. The fractions and counts of variable sources and all sources in each region are listed in Table 2 for  $g < 19$ ,  $g < 20.5$ , and  $g < 22$  flux-limited samples. Notably, in the adopted  $g < 20.5$  flux limit, the fraction of region II sources (dominated by low-redshift quasars) in the variable sample is 63%, or  $\sim 30$  times greater than the fraction of region II sources in the full sample ( $\sim 2\%$ ). The fraction of region IV sources (which include RR Lyrae stars) in the variable sample is also high when compared to the full sample ( $\sim 6$  times higher).

As shown in Table 2, in the  $g = 20.5$  flux-limited sample we find that low-redshift quasars and RR Lyrae stars (i.e., regions II and IV) make up 70% of the variable population while representing only 3% of all sources. Quasars alone account for 63% of the variable population. Stars from the main stellar locus represent 95% of all sources and 25% of the variable sample; about 0.5% of the stars from the locus are variable at the  $\geq 0.05$  mag level.

<sup>26</sup> This list of candidate variable sources is publicly available from [http://www.sdss.org/dr5/products/value\\_added/index.html](http://www.sdss.org/dr5/products/value_added/index.html).

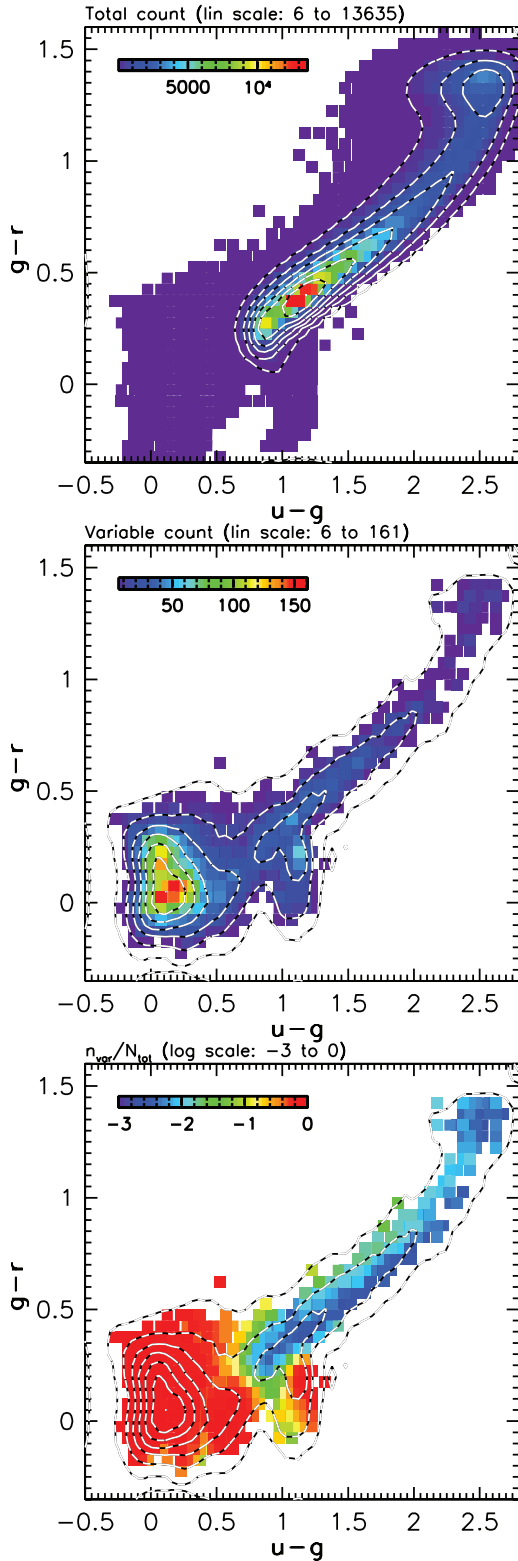


FIG. 3.— Distribution of counts for the full sample (*top*) and candidate variable sample (*middle*), and the ratio of the two counts (*bottom*) in the  $g-r$  vs.  $u-g$  color-color diagram for sources brighter than  $g = 20.5$ , binned in 0.05 mag bins. Contours outline distributions of unbinned counts. Note the remarkable difference between the distribution of all sources and that of the variable sample, which demonstrates that the latter are robustly selected.

TABLE 1  
THE FRACTION OF VARIABLE MAIN STELLAR LOCUS STARS  
AS A FUNCTION OF THE  $s$  COLOR

Bin	% $\sigma(g) \geq 0.05^a$	% Variable <sup>b</sup>	$\langle \sigma(g) \rangle^c$	Counts <sup>d</sup>
$s < -0.02$ .....	3.23	0.36	0.017	46836
$-0.02 < s < 0.02$ .....	2.92	0.28	0.017	136910
$0.02 < s < 0.06$ .....	4.61	1.18	0.019	29106
$s > 0.06$ .....	11.50	4.10	0.027	4547

<sup>a</sup> Fraction of sources with  $\sigma(g) \geq 0.05$  mag.

<sup>b</sup> Fraction of variable sources [selected using  $\sigma(g, r) \geq 0.05$  mag and  $\chi^2(g, r) \geq 3$ ].

<sup>c</sup> Median  $\sigma(g)$ .

<sup>d</sup> Number of sources in the bin.

### 3.3. The Properties of Variable Sources

Various light-curve properties, such as shape and amplitude, are expected to be correlated with stellar types. In this section we study the distribution of the rms scatter in the  $u$  and  $g$  bands and the  $\sigma(g)/\sigma(r)$  ratio as a function of the  $u-g$  and  $g-r$  colors. To emphasize trends, we bin sources and present median values for each bin.

The distribution of the median  $\sigma(u)$  and  $\sigma(g)$  values in the  $g-r$  versus  $u-g$  color-color diagram is shown in Figure 5 (*top panels*). The RR Lyrae stars show larger rms scatter ( $\gtrsim 0.2$ – $0.3$  mag) in the  $u$  and  $g$  bands than the low-redshift quasars or stars from the main stellar locus. The separation of higher amplitude RR Lyrae-type *ab* stars ( $u-g \sim 1.15$ ,  $g-r > 0.15$ ) and lower amplitude type *c* stars ( $u-g \sim 1.15$ ,  $g-r < 0.15$ ) is easily discernible. Quasars also show slightly larger rms scatter in the  $u$  band ( $\sim 0.1$  mag) than in the  $g$  band ( $\sim 0.07$  mag), as discussed by Kinney et al. (1991), Ivezić et al. (2004b), and Vanden Berk et al. (2004). If we define the degree of variability as the rms scatter in the  $g$  band, then, on average, RR Lyrae stars show the greatest variability, followed by quasars and main stellar locus stars.

Another distinctive characteristic of variable sources is the ratio of flux changes in different bandpasses. This property can be used to select different types of variable sources. For example, RR Lyrae stars are bluer when brighter, a behavior used by Ivezić et al. (2000) to select RR Lyrae stars using two-epoch SDSS data. Here we define a new parameter,  $\sigma(g)/\sigma(r)$ , to express the ratio of flux changes in the  $g$  and  $r$  bands, and we study its distribution in the  $g-r$  versus  $u-g$  color-color diagram. In particular, we examine this distribution and its median values for three dominant classes of variable sources: quasars, RR Lyrae stars, and stars from the main stellar locus.

Figure 5 (*bottom left panel*) shows the distribution of median  $\sigma(g)/\sigma(r)$  values as a function of  $u-g$  and  $g-r$  colors. Using Figure 5 we note that, on average,

1. RR Lyrae stars have  $\sigma(g)/\sigma(r) \sim 1.4$ ;
2. main stellar locus stars have  $\sigma(g)/\sigma(r) \sim 1$ ;
3. quasars show a  $\sigma(g)/\sigma(r)$  gradient in the  $g-r$  versus  $u-g$  color-color diagram.

The average value of  $\sigma(g)/\sigma(r) \sim 1.4$  in region IV indicates that RR Lyrae stars dominate the variable source count in this region (Ivezić et al. 2000). While Figure 5 only presents the median values of the rms scatter, Figure 6 shows the correlation between the rms scatter in the  $g$  and  $r$  bands for individual sources. The sources with high rms scatter in the  $g$  band also have high rms scatter in the  $r$  band. Variability in these bands correlates with  $u-g$  color, and for RR Lyrae stars ( $u-g \sim 1$ ) has the form  $\sigma(g) = 1.4\sigma(r)$ . Figure 6 also shows how sources with large  $\chi^2$  ( $\chi^2 > 3$ ) have large rms scatter in the  $g$  and  $r$  bands.

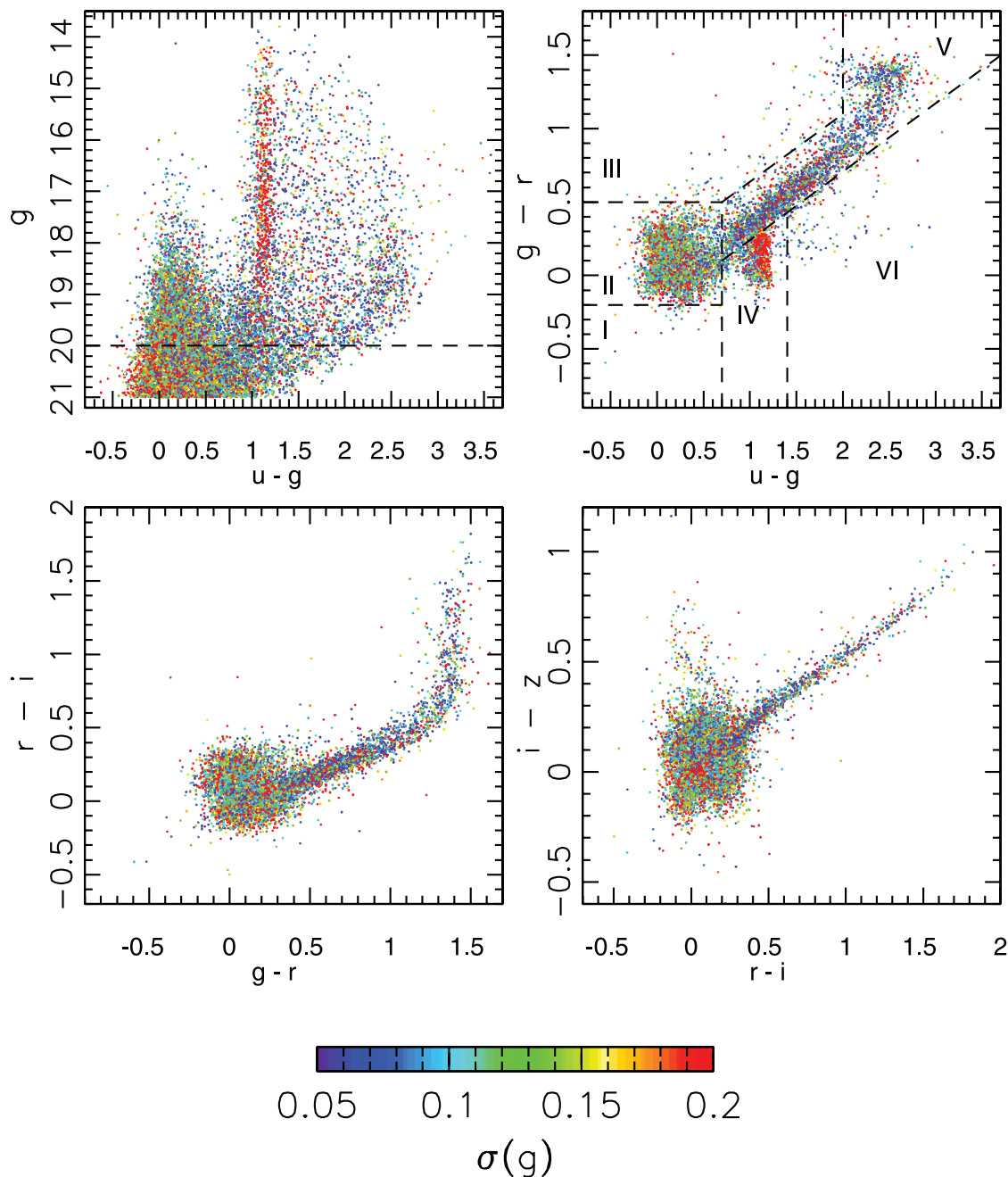


FIG. 4.—Distribution of 18,329 candidate variable sources brighter than  $g = 21$  in representative SDSS color-magnitude and color-color diagrams. Candidate variables are color-coded by their rms scatter in the  $g$  band (for 0.05–0.2, see the legend; for larger than or equal to 0.2, *red*). Only sources brighter than  $g = 20$  are plotted in the color-color diagrams. Note that RR Lyrae stars [ $u - g \sim 1.15$ ,  $\sigma(g) \gtrsim 0.2$  mag; *red dots*] and low-redshift quasars [ $u - g \leq 0.7$ ,  $\sigma(g) \gtrsim 0.1$  mag; *green dots*] stand out as highly variable sources. The regions marked in the top right panel are used for quantitative comparison of the overall and variable source distributions (see Table 2).

The average ratio of  $\sigma(g)/\sigma(r) \sim 1$  (i.e., gray flux variations) for stars in the main stellar locus suggests that variability could be caused by eclipsing systems. The distribution of  $\gamma(g)$  for main stellar locus stars further strengthens this possibility, as discussed in § 3.4.

The gradient in the  $\sigma(g)/\sigma(r)$  ratio observed for low-redshift quasars in the  $g - r$  versus  $u - g$  color-color diagram suggests that the variability correlation between the  $g$  and  $r$  bands is more complex than in the case of RR Lyrae or main stellar locus stars. Wilhite et al. (2005) show that the photometric color changes for quasars depend on the combined effects of continuum changes, emission-line changes, redshift, and the selection of photometric bandpasses. They note that due to the lack of variability of the

lines, measured photometric color is not always bluer in brighter phases but depends on redshift and the filters used. To verify the dependence of broadband photometric variability on redshift, we plot  $\sigma(g)/\sigma(r)$  versus redshift for all spectroscopically confirmed unresolved quasars from Schneider et al. (2005) that are in stripe 82, as shown in Figure 7. We confirm that the broadband photometric variability depends on the redshift, and that the  $\sigma(g)/\sigma(r)$  gradient in the  $g - r$  versus  $u - g$  color-color diagram can be explained by the increase in  $\sigma(g)/\sigma(r)$  from  $\sim 1$  to  $\sim 1.6$  in the 1.0–1.6 redshift range. This effect is due to the Mg II emission line (more stable in flux than the continuum) moving through the  $r$ -band filter over this redshift range. The implied correlation of the  $u - g$  and  $g - r$  colors with redshift is consistent with the discussion by



TABLE 2  
THE DISTRIBUTION OF CANDIDATE VARIABLE SOURCES IN THE  $g - r$  VERSUS  $u - g$  DIAGRAM

REGION <sup>a</sup>	NAME <sup>b</sup>	$g < 19$				$g < 20.5$				$g < 22$			
		% All <sup>c</sup>	% Var. <sup>d</sup>	Var./All <sup>e</sup>	$N_{\text{var}}/N_{\text{all}}^f$	% All <sup>c</sup>	% Var. <sup>d</sup>	Var./All <sup>e</sup>	$N_{\text{var}}/N_{\text{all}}^f$	% All <sup>c</sup>	% Var. <sup>d</sup>	Var./All <sup>e</sup>	$N_{\text{var}}/N_{\text{all}}^f$
I .....	White dwarfs	0.14	0.59	4.25	3.50	0.24	0.40	1.69	3.34	0.28	0.45	1.64	4.51
II .....	Low-redshift QSOs	0.45	30.88	68.83	56.58	1.90	62.90	33.03	65.10	4.07	70.01	17.22	47.30
III .....	dM/WD pairs	0.08	0.53	6.54	5.37	0.83	2.08	2.50	4.92	1.21	3.79	3.13	8.61
IV .....	RR Lyrae stars	1.28	16.81	13.11	10.78	1.33	7.95	5.99	11.81	1.48	6.41	4.33	11.90
V .....	Stellar locus stars	96.27	48.77	0.51	0.42	94.49	25.15	0.27	0.52	91.89	18.33	0.20	0.55
VI .....	High-redshift QSOs	1.78	2.42	1.36	1.12	1.21	1.52	1.26	2.48	1.07	1.01	0.95	2.60
Total count		411667	3384			662195	13051			748067	20553		

<sup>a</sup> These regions are defined in the  $g - r$  vs.  $u - g$  color-color diagram, with their boundaries shown in Fig. 4.

<sup>b</sup> An approximate description of the dominant source type found in the region.

<sup>c</sup> The fraction of all sources in a magnitude-limited sample found in this color region, with the magnitude limits listed on top.

<sup>d</sup> The number of candidate variables from the region, expressed as a fraction of all variable sources.

<sup>e</sup> The ratio of values listed in the two columns immediately preceding.

<sup>f</sup> The number of candidate variables from the region, expressed as a fraction of all sources in that region.

Richards et al. (2002). The lack of noticeable correlation of  $\sigma(g)$  with redshift is due to the combined effects of the dependence of  $\sigma(g)$  on the rest-frame wavelength and time, which cancel out (for a detailed model, see Ivezić et al. 2004b).

### 3.4. Skewness as a Proxy for the Dominant Variability Mechanism

Light-curve skewness, a measure of the light-curve asymmetry, provides additional information on the type of variability. Negatively skewed, asymmetric light curves indicate variable sources that spend more time fainter than  $(m_{\text{min}} + m_{\text{max}})/2$ , where  $m_{\text{min}}$  and  $m_{\text{max}}$  are magnitudes at the minimum and maximum. Type *ab* RR Lyrae stars, for example, have negatively skewed light curves ( $\gamma \sim -0.5$ ; Wils et al. 2006). Positively skewed, asymmetric light curves indicate variable sources that spend more time brighter than  $(m_{\text{min}} + m_{\text{max}})/2$  (e.g., eclipsing systems). Sources with symmetric light curves will have  $\gamma \sim 0$ .

Figure 5 (*bottom right panel*) shows the distribution of the median  $\gamma(g)$  as a function of position in the  $g - r$  versus  $u - g$  color-color diagram. On average, quasars and *c*-type RR Lyrae stars ( $u - g \sim 1.15$ ,  $g - r < 0.15$ ) have  $\gamma(g) \sim 0$ , *ab*-type RR Lyrae stars ( $u - g \sim 1.15$ ,  $g - r > 0.15$ ) have negative skewness [ $\gamma(g) \sim -0.5$ ], and stars in the main stellar locus have positive skewness.

Figure 8 shows the distribution of the light-curve skewness in the *ugr* bands for spectroscopically confirmed unresolved quasars from Schneider et al. (2005), which are in stripe 82; candidate RR Lyrae stars (selection details are discussed in § 4); and main stellar locus stars from our variable sample. Stars in the main stellar locus show a bimodal  $\gamma(g)$  distribution. This distribution suggests at least two, and perhaps more, different populations of variables. Indeed, when spectroscopically confirmed M dwarfs are selected, a third peak appears at  $\gamma(g) \sim 2.5$ , possibly associated with flaring M dwarfs (A. Kowalski et al. 2007, in preparation). A bimodality similar to that in the  $g$  band is also discernible in the  $r$  band, while it is less pronounced in the  $i$  band and not detected in the  $u$  and  $z$  bands (the  $r$  and  $z$  data are not shown). The bimodality in the  $u$  and  $z$  bands is not detected due to high photometric errors in these bands at faint magnitudes. The measurement uncertainty decreases the asymmetry of the light curve, and  $\gamma$  approaches 0. The less pronounced bimodality in the  $i$  band might be due to astrophysical reasons, since the photometric errors in the  $i$  band are comparable to  $g$ - and  $r$ -band photometric errors.

A comparison of the  $u - g$  and  $g - r$  color distributions for variable main stellar locus stars brighter than  $g = 19$  and a subset with highly asymmetric light curves [ $\gamma(g) > 2.5$ ] is shown in Figure 9. The subset with asymmetric light curves has an increased fraction of stars with colors  $u - g \sim 2.5$  and  $g - r \sim 1.4$  that correspond to M stars. This may indicate that M stars have a higher probability of being associated with an eclipsing companion than stars with earlier spectral types. However, the selection effects are probably important, since a companion is easier to detect (due to the low luminosity of M dwarfs). A. Kowalski et al. (2007, in preparation) examine these issues using light-curve data on a sample of spectroscopically confirmed M dwarfs. Finally, quasars have symmetric light curves ( $\gamma \sim 0$ ) and their distribution of skewness does not change between bands.

While the value of the skewness for individual sources may strongly depend on the number of available epochs, the median binned values of skewness are not very sensitive to the number of available epochs. We verify this claim by repeating the analysis presented above on a subsample of sources with at least eight detections in the *ugr* bands. A more conservative cut results in a smaller sample, but general trends and results remain the same.

## 4. THE MILKY WAY HALO STRUCTURE TRACED BY CANDIDATE RR LYRAE STARS

Studies of substructures in the Galactic halo, such as clumps and streams, can constrain the formation history of the Milky Way. Among the best tracers for studying the outer halo are RR Lyrae stars, because they are nearly standard candles, are sufficiently bright to be detected at large distances (5–100 kpc for  $14 < r < 20.7$ ), and are sufficiently numerous to trace the halo substructure with a high spatial resolution. The General Catalog of Variable Stars (GCVS; Kholopov et al. 1988) lists<sup>27</sup> RR Lyrae stars as RR Lyrae type *ab* (RR*ab*) and type *c* (RR*c*) stars. RR*ab* stars have asymmetric light curves, periods from 0.3 to 1.2 days, and amplitudes from  $V \sim 0.5$  to 2. RR*c* stars have nearly symmetric, sometimes sinusoidal, light curves with periods from 0.2 to 0.5 days and amplitudes not greater than  $V \sim 0.8$ . In this work we assume  $M_V = 0.7$  as the absolute  $V$ -band magnitude of RR*ab* and RR*c* stars. A comprehensive review of RR Lyrae stars can be found in Smith (1995).

<sup>27</sup> A list of GCVS variability types can be found at <http://www.sai.msu.ru/groups/cluster/gcvs/gcvs/iii/vartype.txt>.

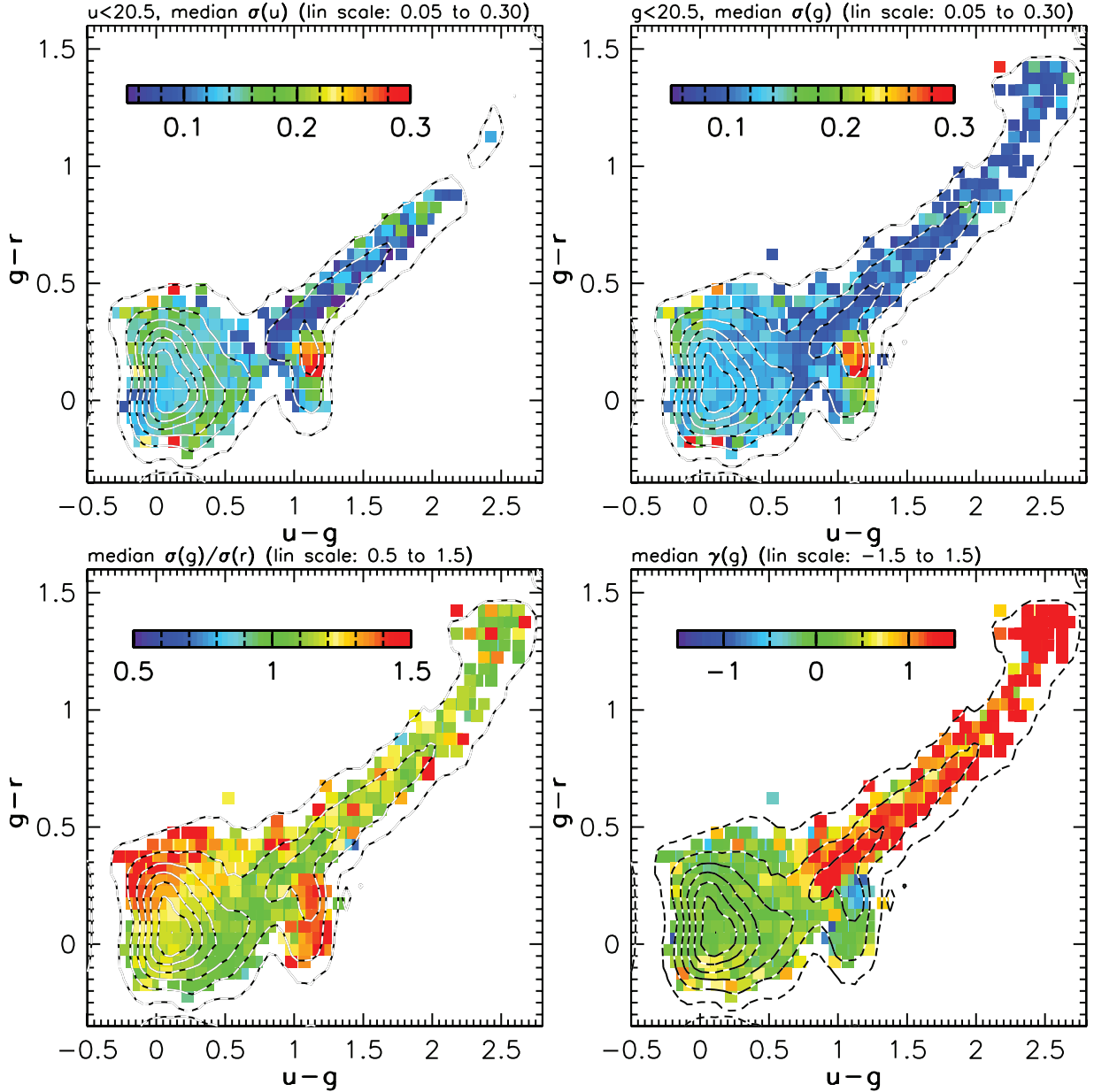


FIG. 5.—Distribution of the rms scatter  $\sigma(u)$  (top left), rms scatter  $\sigma(g)$  (top right),  $\sigma(g)/\sigma(r)$  ratio (bottom left), and  $\gamma(g)$  (bottom right) for the variable sample in the  $g-r$  vs.  $u-g$  color-color diagram. Sources are binned in 0.05 mag wide bins, and the median values are color-coded. Color ranges are given at the top of each panel, from blue to red, where green is in the midrange. Values outside the range saturate in blue or red. Contours outline the count distributions on a linear scale in steps of 15%. The flux limit is  $g < 20.5$ , with an additional  $u < 20.5$  limit in the top left panel. Top panels: Note how higher amplitude RRab ( $u-g \sim 1.15$ ,  $g-r > 0.15$ ) and lower amplitude RRC stars ( $u-g \sim 1.15$ ,  $g-r < 0.15$ ) separate in these plots. Bottom left: On average, RR Lyrae stars have  $\sigma(g)/\sigma(r) \sim 1.4$ , main stellar locus stars have  $\sigma(g)/\sigma(r) \sim 1$ , and low-redshift quasars show a gradient of  $\sigma(g)/\sigma(r)$  values. Bottom right: On average, quasars and RRC stars have  $\gamma(g) \sim 0$ , RRab stars have negative skewness, and stars in the main stellar locus have positive skewness.

In this section we fine-tune criteria for selecting candidate RR Lyrae stars and estimate the selection completeness and efficiency. Using selected candidate RR Lyrae stars, we recover a known halo clump associated with the Sgr dwarf tidal stream and find several new halo substructures.

#### 4.1. Criteria for Selecting RR Lyrae Stars

Figures 3–5 show that RR Lyrae stars occupy a well-defined region (region IV) in the  $g-r$  versus  $u-g$  color-color diagram, and Figure 6 shows how RR Lyrae stars follow the  $\sigma(g) = 1.4\sigma(r)$  relation. Motivated by these results, we introduce color and  $\sigma(g)/\sigma(r)$  cuts to specifically select candidate RR Lyrae stars

from the variable sample and study their distribution in the rms scatter–color–light-curve skewness parameter space.

RR Lyrae stars have distinctive colors and can be selected with the following criteria (Ivezić et al. 2005):

$$0.98 < u - g < 1.30, \quad (9)$$

$$-0.05 < D_{ug} < 0.35, \quad (10)$$

$$0.06 < D_{gr} < 0.55, \quad (11)$$

$$-0.15 < r - i < 0.22, \quad (12)$$

$$-0.21 < i - z < 0.25, \quad (13)$$

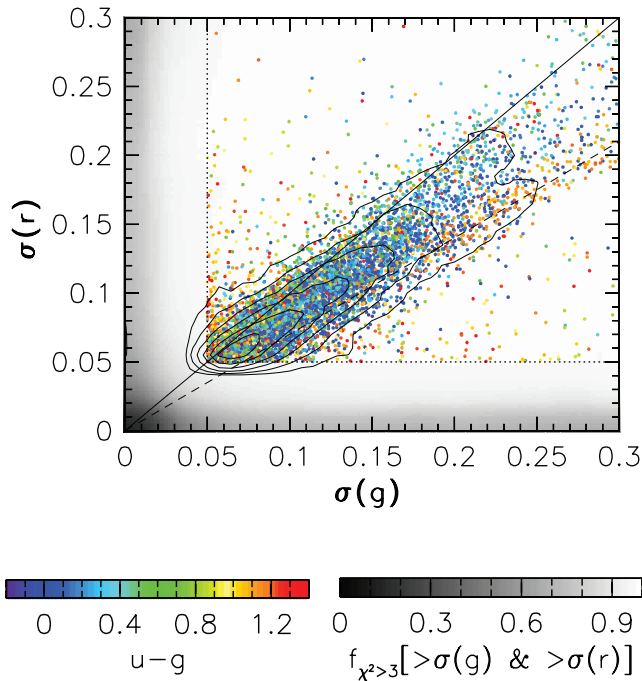


FIG. 6.—Distribution of candidate variable sources in the  $g < 20.5$  flux-limited sample. This is shown by linearly spaced contours and by symbols color-coded with the  $u - g$  color for sources with  $\sigma(g) \geq 0.05$  and  $\sigma(r) \geq 0.05$  mag. The dotted lines show the adopted  $\sigma(g, r)$  selection cut. The solid line shows  $\sigma(g) = \sigma(r)$ , while the dashed line shows the  $\sigma(g) = 1.4\sigma(r)$  relation representative of RR Lyrae stars. Note that sources following the  $\sigma(g) = 1.4\sigma(r)$  relation tend to have  $u - g \sim 1$ , as expected for RR Lyrae stars. The gray-scale background shows the fraction of  $\chi^2(g, r) \geq 3$  that also has  $\sigma(g) \geq x$  and  $\sigma(r) \geq y$  and demonstrates that large  $\chi^2$  sources also have large  $\sigma$ .

where

$$D_{ug} = (u - g) + 0.67(g - r) - 1.07 \quad (14)$$

and

$$D_{gr} = 0.45(u - g) - (g - r) - 0.12. \quad (15)$$

We apply these cuts to our sample of candidate variables and select 846 sources. It is implied by Ivezić et al. (2005) that RR Lyrae stars should always stay within these color boundaries, even though their colors change as a function of phase. Their distribution in the  $g - r$  versus  $u - g$  color-color diagram and rms scatter in the  $g$  band are shown in Figure 10 (*top left panel*). The distribution of sources in the RR Lyrae region is inhomogeneous. Sources with large rms scatter in the  $g$  band ( $\gtrsim 0.2$  mag) are centered around  $u - g \sim 1.15$  and are separated by  $g - r \sim 0.12$  into two groups. A comparison with Figure 3 from Ivezić et al. (2005) suggests that these large rms scatter sources might be RRab ( $g - r > 0.12$ ) and RRc ( $g - r < 0.12$ ) stars. Small rms scatter sources ( $\lesssim 0.1$  mag) have a fairly uniform distribution and are slightly bluer, with  $u - g \lesssim 1.1$ .

The distribution of sources from the RR Lyrae region in the  $\sigma(r)$  versus  $\sigma(g)$  diagram is presented in Figure 10 (*top right panel*). The majority of large rms scatter sources follow the  $\sigma(g) = 1.4\sigma(r)$  relation, as expected for RR Lyrae stars. Since RR Lyrae stars are bluer when brighter or, equivalently, have greater rms scatter in the  $g$  band than in the  $r$  band, we require  $1 < \sigma(g)/\sigma(r) \leq 2.5$  and select 683 candidate RR Lyrae stars.

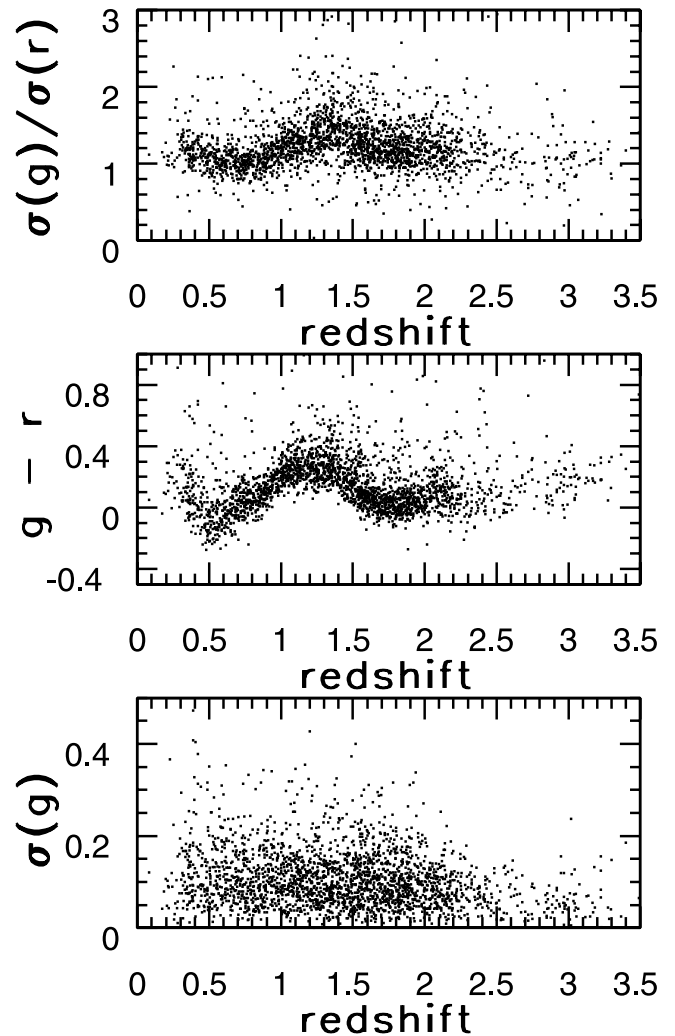


FIG. 7.—Dependence of  $\sigma(g)/\sigma(r)$  (*top*),  $g - r$  (*middle*), and  $\sigma(g)$  (*bottom*) on redshift for a sample of spectroscopically confirmed unresolved quasars from Schneider et al. (2005). The  $\sigma(g)/\sigma(r)$  gradient shown in Fig. 5 (*bottom left panel*) can be explained by the local maximum of  $\sigma(g)/\sigma(r)$  in the 1.0–1.6 redshift range.

A comparison of  $u - g$  color distributions for candidate RR Lyrae stars and of sources with RR Lyrae colors but not tagged as RR Lyrae stars, presented in Figure 10 (*bottom left panel*), demonstrates the robustness of the RR Lyrae selection. The two distributions are very different (the probability that they are the same is  $10^{-4}$ , as given by the Kolmogorov-Smirnov test), with the candidate RR Lyrae distribution peaking at  $u - g \sim 1.15$ , as expected for RR Lyrae stars.

One property that distinguishes RRab from RRc stars is the shape (or skewness) of their light curves (in addition to light-curve amplitude and period). RRab stars have asymmetric light curves, while RRc light curves are symmetric. In Figure 10 (*top left panel*) we noted that  $g - r \sim 0.12$  seemingly separates high rms scatter sources into two groups. If  $g - r \sim 0.12$  is the boundary between the RRab and RRc stars, then the same boundary should show up in the distribution of light-curve skewness as a function of  $g - r$  color. As shown in Figure 10 (*bottom right panel*), this is indeed the case. On average, sources with  $g - r < 0.12$  have  $\gamma(g) \sim 0$  (symmetric light curves), as do RRc stars, while  $g - r > 0.12$  sources have  $\gamma(g) \sim -0.5$  (asymmetric light curves) typical of RRab stars.

We show in § 4.2 that candidate RR Lyrae stars with  $\gamma(g) > 1$  are contaminated by eclipsing variables. Therefore, to reduce the

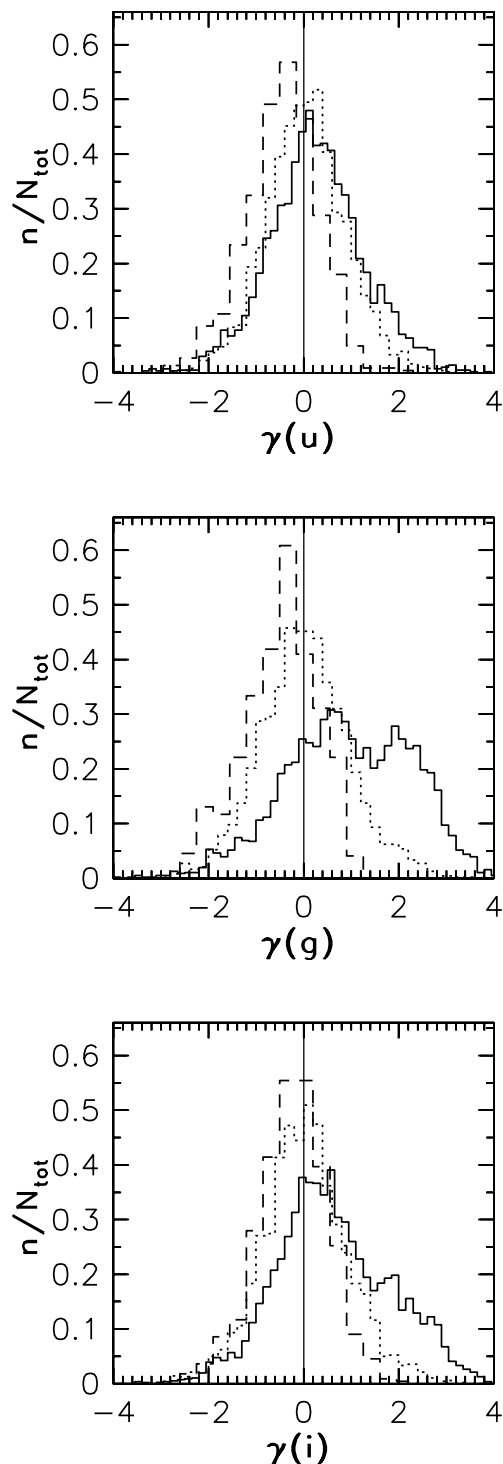


FIG. 8.—Light-curve skewness distribution in the  $ugi$  bands for spectroscopically confirmed unresolved quasars (dotted line), candidate RR Lyrae stars (dashed line), and variable main stellar locus stars (solid line, region V; see Fig. 4 for the definition). The distribution of the skewness in the  $r$  band is similar to the  $g$ -band distribution, and the distribution of skewness in the  $z$  band is similar to the  $u$ -band distribution (therefore, the  $r$  and  $z$  data are not shown). Stars in the main stellar locus show bimodality in  $\gamma(g)$ , suggesting at least two, and perhaps more, different populations of variables. Similar bimodality is also discernible in the  $r$  band, while it is less pronounced in the  $i$  band and not detected in the  $u$  and  $z$  bands (due to higher photometric errors). Quasars have symmetric light curves ( $\gamma \sim 0$ ), and their distribution of skewness does not change between bands. [See the electronic edition of the *Journal* for a color version of this figure.]

contamination by eclipsing variables, we also require  $\gamma(g) \leq 1$  and select 634 sources as our final sample of candidate RR Lyrae stars.

#### 4.2. Completeness and Efficiency

The selection completeness, defined as the fraction of recovered RR Lyrae stars, will depend on the color cuts, the  $\sigma(g, r)$  cutoff, and the number of observations. The color cuts (eqs. [9]–[15]) applied in § 4.1 were chosen to minimize contamination by sources other than RR Lyrae stars while maintaining an almost 100% completeness (Ivezić et al. 2005). With the  $\sigma(g, r)$  cutoff at 0.05 mag (small compared to the  $\sim 1$  mag typical peak-to-peak amplitudes of RR Lyrae stars) and a fairly large number of observations per source (median of 10), we estimate the RR Lyrae selection completeness to be  $\geq 95\%$  (see the Appendix in Ivezić et al. 2000).

To determine the selection efficiency, defined as the fraction of true RR Lyrae stars in the RR Lyrae candidate sample, we positionally match 683 candidate RR Lyrae stars selected by  $1 < \sigma(g)/\sigma(r) \leq 2.5$  to a sample of RR Lyrae sources selected from the SDSS Light-Motion-Curve Catalog (LMCC; D. Bramich et al. 2007, in preparation). This catalog covers a slightly smaller region of the sky ( $20^{\text{h}}42^{\text{m}} < \alpha_{\text{J2000.0}} < 03^{\text{h}}16^{\text{m}}$ ,  $-1.26^\circ < \delta_{\text{J2000.0}} < +1.26^\circ$ ) than that discussed here but includes more densely sampled SDSS-II observations that allow the construction of light curves. We match 613 candidates, while 70 candidate RR Lyrae stars from our sample are not in the LMCC footprint. Following the classification based on phased light curves by N. De Lee et al. (2007, in preparation) we find that 71% of  $1 < \sigma(g)/\sigma(r) \leq 2.5$ ,  $\gamma(g) \leq 1$  sources in our candidate RR Lyrae sample are classified as RRab and RRc, 28% are classified as variable non-RR Lyrae stars, and only 1% of sources in this sample are classified as spurious, nonvariable sources. While we do not know exactly the completeness of the N. De Lee et al. (2007, in preparation) samples (efficiency is about 100%, as verified by the visual inspection of light curves), we speculate that it is likely very high, given more densely sampled observations available in the SDSS-II Supernova Survey. The most significant contamination in our candidate RR Lyrae sample comes from a population of variable sources bluer than  $u - g \sim 1.1$  (Fig. 11, dotted line in bottom left panel), possibly Population II  $\delta$  Scuti stars, also known as SX Phoenixis stars (Hoffmeister et al. 1985).

The top left and the bottom right panels in Figure 11 show that RRab- and RRc-dominated regions are separated by  $g - r \sim 0.12$ , as already hinted in Figure 10. Also, variable non-RR Lyrae sources with  $\gamma(g) > 1$  are classified by N. De Lee et al. (2007, in preparation) as eclipsing variables, justifying our  $\gamma(g) \leq 1$  cut.

To summarize, using color criteria and criteria based on  $\sigma(g)$ ,  $\sigma(r)$ , and  $\gamma(g)$ , RR Lyrae stars are selected with  $\geq 95\%$  completeness and  $\sim 70\%$  efficiency.

#### 4.3. The Spatial Distribution of Candidate RR Lyrae Stars

Using the selection criteria from § 4.1 we isolate 634 RR Lyrae candidates. The magnitude-position diagram for these candidates within  $2.5^\circ$  of the celestial equator is shown in Figure 12.

As discussed by Ivezić et al. (2005), an advantage of the data representation utilized in Figure 12 (magnitude–right ascension diagram) is its simplicity; only “raw” data are shown, without any postprocessing. However, the magnitude scale is logarithmic, and thus the spatial extent of structures is heavily distorted. In order to avoid these shortcomings, we have applied a Bayesian method for estimating continuous spatial density distribution developed by Ivezić et al. (2005, their Appendix B). The resulting density map is shown in Figure 13 (right). The advantage of that representation

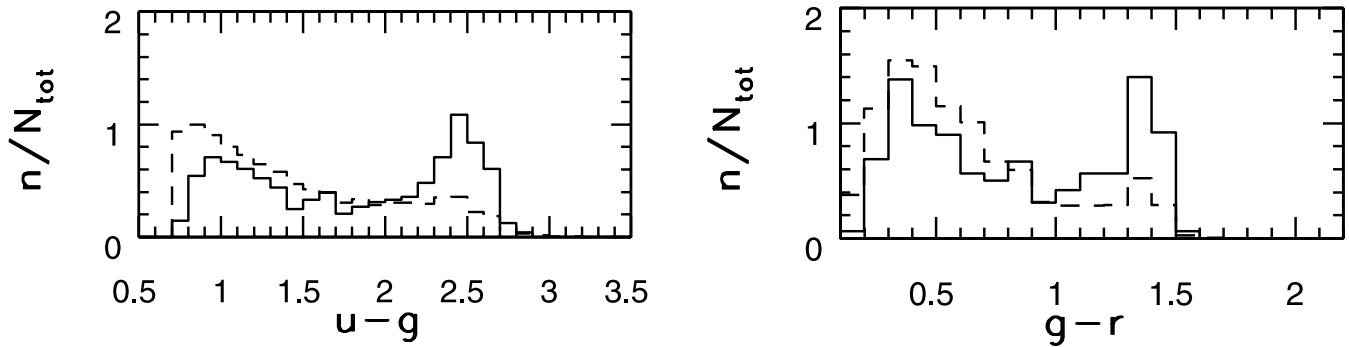


FIG. 9.—Comparison of the  $u - g$  (left) and  $g - r$  (right) color distributions for variable main stellar locus stars brighter than  $g = 19$  (dashed lines) and a subset with highly asymmetric light curves [ $\gamma(g) > 2.5$ ; solid lines]. The subset with highly asymmetric light curves has an increased fraction of stars with colors  $u - g \sim 2.5$  and  $g - r \sim 1.5$ , characteristic of M stars. [See the electronic edition of the Journal for a color version of this figure.]

is that it better conveys the significance of various local overdensities. For comparison, we also show a map of the northern part of the equatorial strip constructed using two-epoch data discussed by Ivezić et al. (2000).

We detect several new halo substructures at  $\geq 3\sigma$  significance (compared to expected Poissonian fluctuations) and present their

approximate locations and properties in Table 3. In order to assess the statistical significance of the clumps seen in Figure 13, we have performed a detailed analysis using random samples of the same size as the candidate RR Lyrae sample. The samples were constructed using flat distributions in magnitude and right ascension, which approximately correspond to a  $1/R^3$  RR Lyrae volume

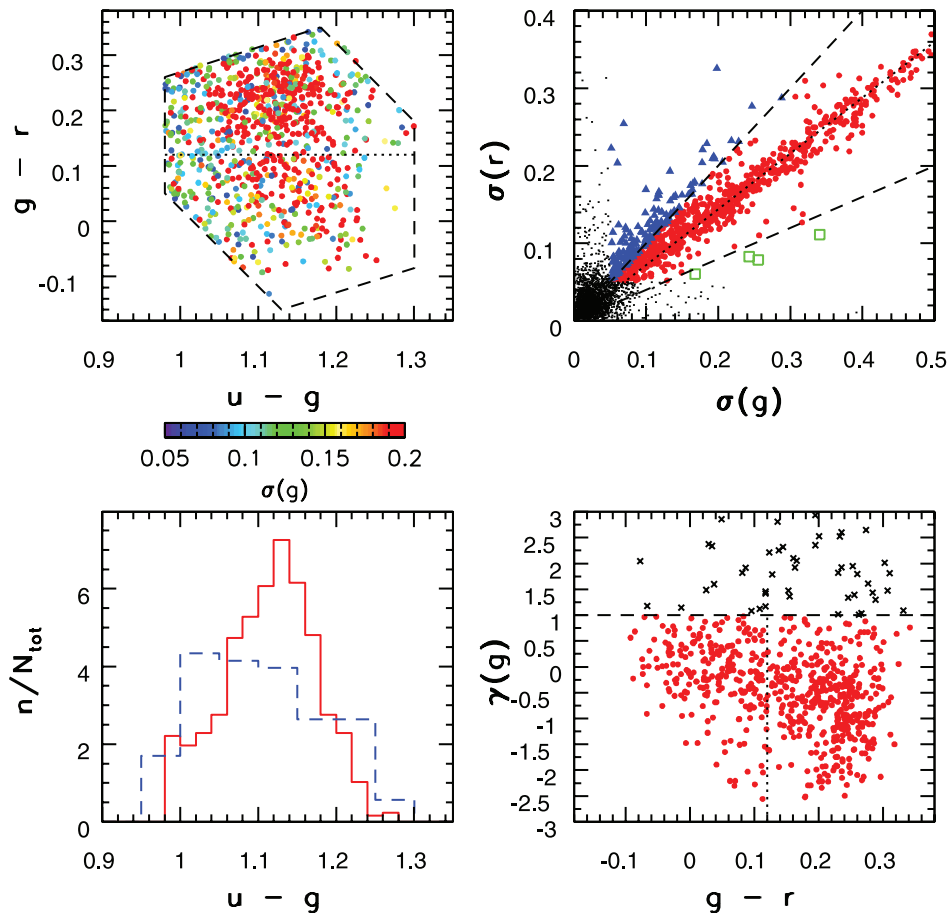


FIG. 10.—Top left: Distribution of 846 candidate variable sources from the RR Lyrae region (dashed lines; see Fig. 3 in Ivezić et al. 2005) in the  $g - r$  vs.  $u - g$  color-color diagram. The symbols mark the time-averaged values and are color-coded by  $\sigma(g)$  (0.05–0.2; blue to red). The dotted line shows the boundary between the RRab- and RRc-dominated regions. Top right: Sources from the top left panel divided into three groups according to their  $\sigma(g)/\sigma(r)$  values: candidate RR Lyrae stars with  $1 < \sigma(g)/\sigma(r) \leq 2.5$  (large dots), sources with  $\sigma(g)/\sigma(r) \leq 1$  (triangles), and sources with  $\sigma(g)/\sigma(r) > 2.5$  (squares). Small dots show sources with RR Lyrae colors that fail the variability criteria. The dashed lines show the  $\sigma(g) = \sigma(r)$  and  $\sigma(g) = 2.5\sigma(r)$  relations, while the dotted line shows the  $\sigma(g) = 1.4\sigma(r)$  relation. Bottom left: Comparison of the  $u - g$  color distributions for candidate RR Lyrae stars (solid line) and sources with RR Lyrae colors but not tagged as RR Lyrae stars (dashed line). Bottom right: Dependence of  $\gamma(g)$  on the  $g - r$  color for candidate RR Lyrae stars. The boundary  $g - r = 0.12$  (dotted line) separates candidate RR Lyrae stars into those with asymmetric [ $\gamma(g) \sim -0.5$ ] and symmetric [ $\gamma(g) \sim 0$ ] light curves, corresponding to RRab and RRc stars, respectively. The condition  $\gamma(g) \leq 1$  (dashed line) is used to reduce the contamination of the RR Lyrae sample by eclipsing variables.

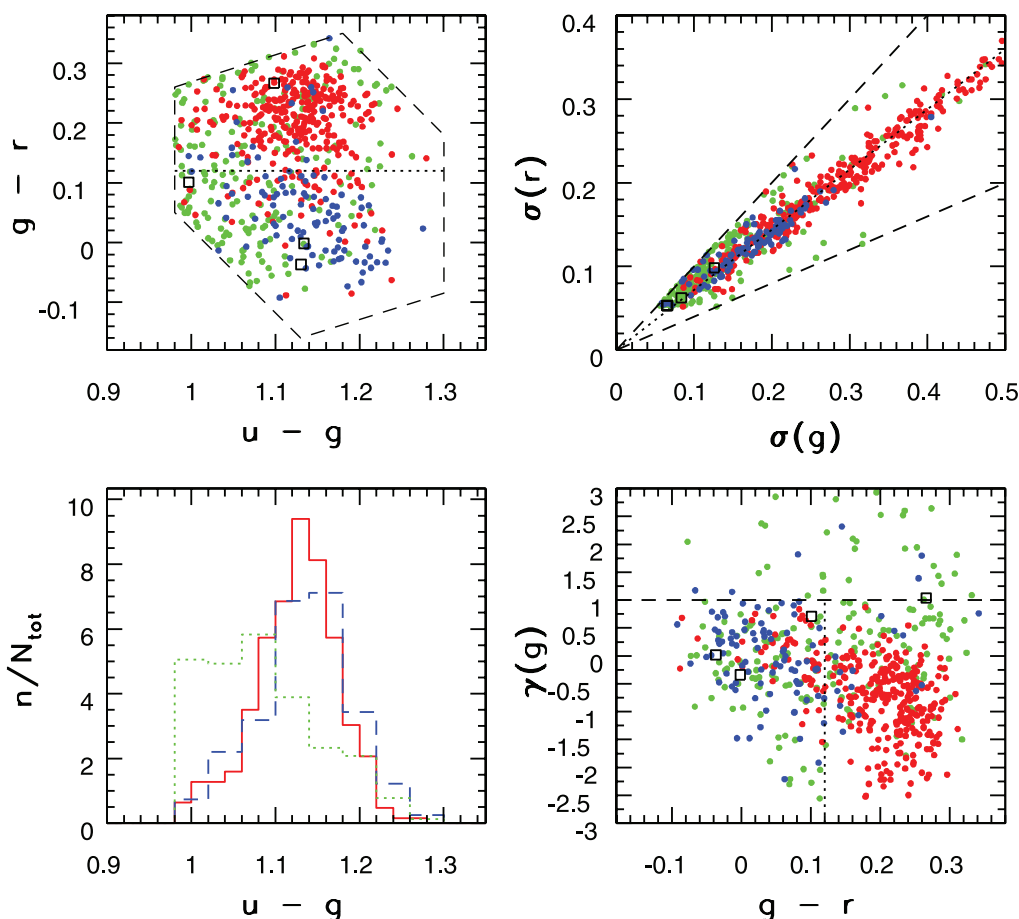


FIG. 11.—Distribution of candidate RR Lyrae stars selected with  $1 < \sigma(g)/\sigma(r) \leq 2.5$  and classified by N. De Lee et al. (2007, in preparation) shown in diagrams similar to Fig. 10. Symbols show RRab stars (red dots), RRC stars (blue dots), variable non-RR Lyrae stars (green dots), and nonvariable sources (open squares; only four sources). A comparison of the  $u - g$  color distribution for RRab (solid line), RRC (dashed line), and variable non-RR Lyrae stars (dotted line) is shown in the bottom left panel.

density distribution, where  $R$  is the distance from the Galactic center. We find that the contamination by non-RR Lyrae stars is not an issue because this effect tends to minimize the structure, rather than enhance it. The random samples often have clumpy features due to Poisson fluctuations. By comparing the histograms of density distributions for random samples and the observed candidate RR Lyrae sample, we conclude that clumps I and K are consistent with random fluctuations, and one or two of clumps D, F,

H, and L may also be caused by such fluctuations. The remaining seven clumps (A, B, C, E, G, J, and M) are inconsistent with random fluctuations and likely represent real halo substructure. Stars associated with these clumps account for about 50% of the sample. This estimate for the fraction of halo stars that are associated with substructure is in good agreement with estimates based on main-sequence stars in the inner halo (out to about 30 kpc) by Jurić et al. (2007) and Bell et al. (2007).

The most distant clump (M) is 106 kpc from the Galactic center. The strongest clump in the left wedge belongs to the Sgr dwarf tidal stream, as does clump C in the right wedge (Ivezić et al. 2003). Similarly to the behavior of main-sequence stars discussed by Bell et al. (2007), the apparent “clumpiness” of the candidate RR Lyrae distribution increases with increasing radius, as predicted by CDM simulations (Bullock et al. 2001). A detailed comparison of their models with the data presented here will be discussed elsewhere (B. Sesar et al. 2007, in preparation).

## 5. ARE ALL QUASARS VARIABLE?

The optical continuum variability of quasars has been recognized since their first optical identification (Matthews & Sandage 1963), and it has been proposed and utilized as an efficient method for their discovery (van den Bergh et al. 1973; Hawkins 1983; Koo et al. 1986; Hawkins & Veron 1995; Rengstorff et al. 2004). The observed characteristics of the variability of quasars are frequently used to constrain the origin of their emission (e.g., Kawaguchi et al. 1998 and references therein; Martini & Schneider

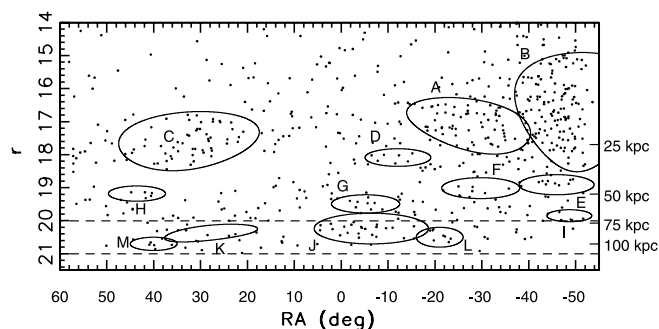


FIG. 12.—Magnitude-position distribution of 634 stripe 82 RR Lyrae candidates within  $-55^\circ < \text{R.A.} < 60^\circ$  and  $|\text{decl.}| \leq 1.27^\circ$ . Approximate distance (shown on the right y-axis) is calculated assuming  $M_r = 0.7$  mag for RR Lyrae stars. Dashed lines show where sample completeness decreases from approximately 99% to 60% due to the  $\chi^2$  cut (see Fig. 2, bottom right panel). Closed curves are remapped ellipses and circles from Fig. 13 that mark halo substructure. [See the electronic edition of the *Journal* for a color version of this figure.]



TABLE 3  
APPROXIMATE LOCATIONS AND PROPERTIES OF DETECTED OVERDENSITIES

Label <sup>a</sup>	$N^b$	$\langle R.A. \rangle^c$	$\langle d \rangle^d$	$\langle r \rangle^e$	$\langle u - g \rangle^f$	$\langle g - r \rangle^g$
A.....	84	330.95	21	17.02	1.14	0.18
B.....	144	309.47	22	16.76	1.12	0.16
C.....	54	33.69	25	17.61	1.13	0.20
D.....	8	347.91	29	18.02	1.14	0.23
E.....	11	314.06	43	18.84	1.09	0.20
F.....	11	330.26	48	19.16	1.07	0.20
G.....	10	354.81	55	19.46	1.10	0.22
H.....	7	43.57	57	19.32	1.05	0.04
I.....	4	311.34	72	19.98	1.08	0.11
J.....	26	353.58	81	20.21	1.11	0.20
K.....	8	28.39	84	20.35	1.10	0.20
L.....	3	339.01	92	20.45	1.06	0.16
M.....	5	39.45	102	20.73	1.07	0.11

<sup>a</sup> The overdensity's label from Fig. 13.

<sup>b</sup> Number of candidate RR Lyrae stars in the overdensity.

<sup>c</sup> Median right ascension.

<sup>d</sup> Median heliocentric distance (in kpc), computed using eq. (3) from Ivezić et al. (2005) and  $M_V = 0.7$  mag as the absolute magnitude of RR Lyrae stars in the  $V$  band.

<sup>e</sup> Median  $r$ -band magnitude.

<sup>f</sup> Median  $u - g$  color.

<sup>g</sup> Median  $g - r$  color.

2003; Pereyra et al. 2006). Recently, significant progress in the description of quasar variability has been made by employing the SDSS data (de Vries et al. 2003, 2005; Ivezić et al. 2004b; Vanden Berk et al. 2004; Sesar et al. 2006). Here we expand these studies by quantifying the efficiency of quasar discovery using variability.

A preliminary comparison of color- and variability-based methods for selecting quasars using SDSS data was presented by Ivezić et al. (2004c). They found that 47% of spectroscopically confirmed unresolved quasars with UV excess have a  $g$ -band magnitude difference between two observations obtained 2 yr apart larger than 0.15 mag. We can improve on their analysis because now there are significantly more observations obtained over a longer time period. Since quasars vary erratically and the rms scatter of their variability (the so-called structure function) increases with time (e.g., Vanden Berk et al. 2004 and references therein), the variability selection completeness is expected to be higher than  $\sim 50\%$ , obtained by Ivezić et al. (2004c).

First, although the adopted variability selection criteria discussed above are fairly conservative, we find that at least 63% of low-redshift quasars are variable at the  $\geq 0.05$  mag level (simultaneously in the  $g$  and  $r$  bands over observer's timescales of several years) in the  $g < 20.5$  flux-limited sample. Second, even this estimate is only a lower limit; given the spectroscopic confirmation for a large flux-limited sample of quasars, it is possible to relax the adopted variability selection cutoff without prohibitive contamination by nonvariable sources.

There are 2492 unresolved quasars in the catalog of spectroscopically confirmed SDSS quasars (Schneider et al. 2005) from stripe 82. The fraction of these objects that vary more than  $\sigma$  in the  $g$  and  $r$  bands, as a function of  $\sigma$ , is shown in Figure 14. We also show the analogous fraction for stars from the stellar locus. About 93% of quasars vary with  $\sigma > 0.03$  mag. For a small fraction of these objects the measured rms scatter is due to photometric noise, and the stellar data limit this fraction to at most 3%. Conservatively assuming that none of these 3% of stars are intrinsically

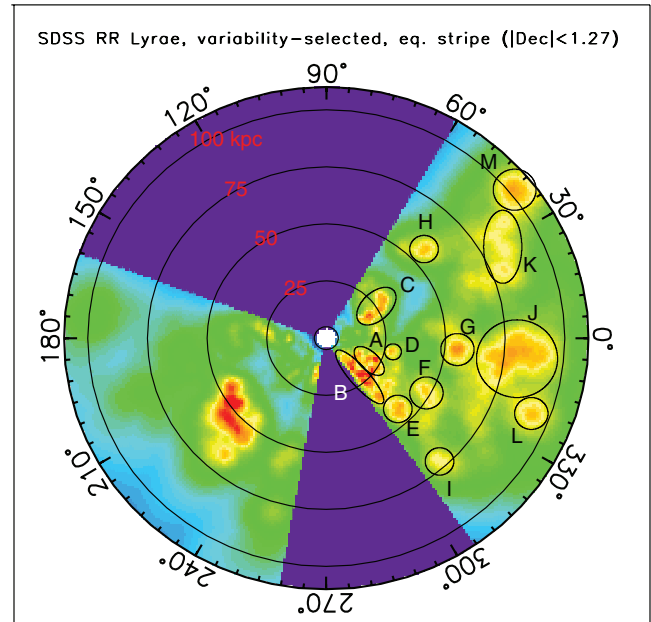
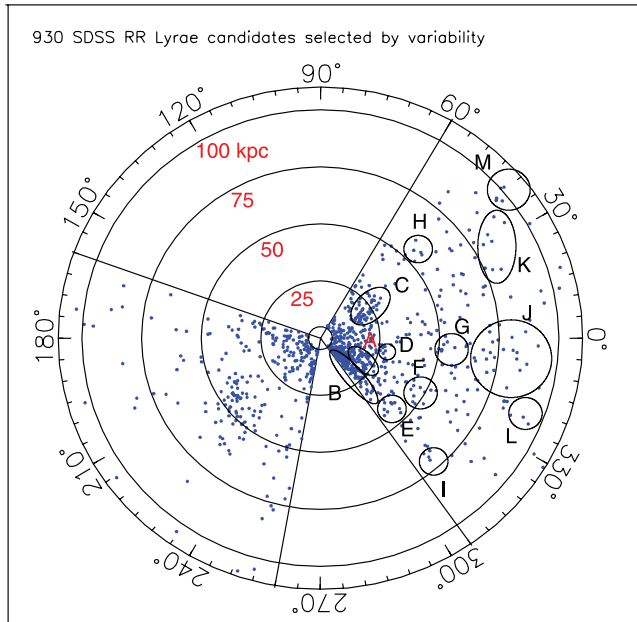


FIG. 13.—*Left*: Spatial distribution of candidate RR Lyrae stars discovered by SDSS along the celestial equator. Distance is calculated assuming eq. (3) from Ivezić et al. (2005) and  $M_V = 0.7$  mag as the absolute magnitude of RR Lyrae stars in the  $V$  band. The right wedge corresponds to candidate RR Lyrae stars selected in this work (634 candidates, shown in Fig. 12), and the left wedge is based on the sample from Ivezić et al. (2000; 296 candidates). *Right*: Number density distribution of candidate RR Lyrae stars shown in the left panel, computed using an adaptive Bayesian density estimator developed by Ivezić et al. (2005). The color scheme represents the number density multiplied by the cube of the galactocentric radius and displayed on a logarithmic scale with a dynamic range of 300 (light blue to red). Green corresponds to the mean density; both wedges with the data would have this color if the halo number density distribution followed a perfectly smooth  $r^{-3}$  power law. Purple marks the regions with no data. The yellow/orange regions are formally about  $3\sigma$  significant overdensities, and red regions have an even higher significance (using only the count variance). A comparison of this map with those generated using random samples of the same size suggests that clumps I and K are consistent with random fluctuations; one or two of clumps D, F, H, and L may also be caused by such fluctuations; and it is highly likely that the remaining seven clumps (A, B, C, E, G, J, and M) represent real halo substructure (they account for about 50% of the sample in the right wedge). The strongest clump in the left wedge belongs to the Sgr dwarf tidal stream, as does clump C in the right wedge (Ivezić et al. 2003). An approximate location and properties of labeled overdensities are listed in Table 3. The Ivezić et al. (2000) sample is based on only two epochs and thus has a much lower completeness ( $\sim 56\%$ ), resulting in a lower density contrast.

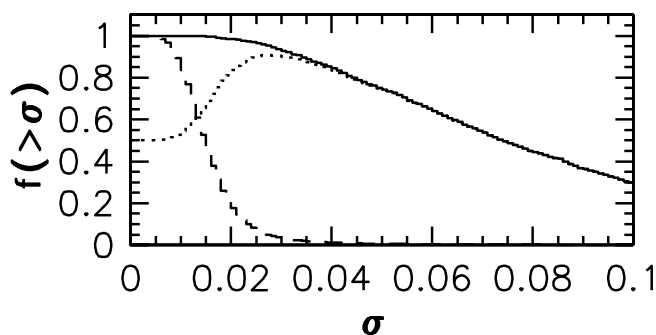


FIG. 14.—Fraction of spectroscopically confirmed unresolved QSOs ( $f_{\text{QSO}}$ , solid line) and fraction of sources from the stellar locus ( $f_{\text{loc}}$ , dashed line) brighter than  $g = 19.5$  and  $r = 19.5$  that have rms scatter larger than  $\sigma$  in the  $g$  and  $r$  bands. The ratio  $f_{\text{QSO}}/(1 + f_{\text{loc}})$  (dotted line), which corresponds to the implied fraction of variable QSOs, peaks at a level of 90% for  $\sigma = 0.03$  mag. [See the electronic edition of the Journal for a color version of this figure.]

variable, we estimate that *at least 90% of quasars are variable at the 0.03 mag level on timescales up to several years.*

## 6. IMPLICATIONS FOR SURVEYS SUCH AS LSST

The LSST is a proposed imaging survey that aims to obtain repeated multiband imaging to faint limiting magnitudes over a large fraction of the sky. The LSST Science Requirements Document<sup>28</sup> calls for  $\sim 1000$  repeated observations of a solid angle of  $\sim 20,000$  deg<sup>2</sup> distributed over the six *ugrizY* photometric bandpasses and over 10 yr. The results presented here can be extrapolated to estimate the lower limit on the number of variable sources that the LSST would discover.

The single-epoch LSST images will have a  $5\sigma$  detection limit<sup>29</sup> at  $r \sim 24.7$ . Hence, 2% accurate photometry, comparable to the subsample with  $g < 20.5$  discussed here, will be available for stars with  $r \lesssim 22$ . The USNO-B catalog (Monet et al. 2003) shows that there are about  $10^9$  stars with  $r < 21$  across the entire sky. About half of these stars are in the parts of the sky to be surveyed by the LSST. The simulations based on contemporary Milky Way models, such as those developed by Robin et al. (2003) and Jurić et al. (2007), predict that there are about twice as many stars with  $r < 22$  than with  $r < 21$  across the whole sky. Hence, it is expected that the LSST will detect about a billion stars with  $r < 22$ . This estimate is uncertain to within a factor of 2 or so due to unknown details in the spatial distribution of dust in the Galactic plane and toward the Galactic center.

We found that at least 0.5% of stars from the main stellar locus can be detected as variable with photometry accurate to  $\sim 2\%$ . This is only a lower limit because a much larger number of LSST observations obtained over a longer time span than the SDSS data discussed here would increase this fraction. Ongoing LSST simulations suggest an increase by a factor of 10 for stars with  $r < 24.5$ . Hence, our results imply that *the LSST will discover at least 50 million variable stars* (without accounting for the fact that stellar counts greatly increase closer to the Galactic plane). Unlike the SDSS sample, where RR Lyrae stars account for  $\sim 25\%$  of all variable stars, the number of RR Lyrae stars in the LSST sample will be negligible compared to other types of variable stars.

As estimated by Jurić et al. (2007) using deeper co-added SDSS photometry, there are about  $100$  deg<sup>-2</sup> low-redshift quasars with  $r < 22$  (see also Beck-Winchatz & Anderson 2007 and references therein). Therefore, with a sky coverage of  $\sim 20,000$  deg<sup>2</sup>,

the LSST will obtain well-sampled, accurate, multicolor light curves for  $\sim 2$  million low-redshift quasars. Even at the redshift limit of  $\sim 2$ , this sample will be complete to  $M_r \sim -24$ , that is, almost to the formal quasar luminosity cutoff, and will represent an unprecedented sample for studying quasar physics.

## 7. CONCLUSIONS AND DISCUSSION

We have designed and tested algorithms for selecting candidate variable sources from a catalog based on multiple SDSS imaging observations. Using a sample of 13,051 selected candidate variable sources in the adopted  $g < 20.5$  flux-limited sample, we find that at least 2% of unresolved optical sources at high Galactic latitudes ( $b < -20^\circ$ ) appear variable at the  $\geq 0.05$  mag level simultaneously in the  $g$  and  $r$  bands. A similar fraction of variable sources ( $\sim 1\%$ ) was also found by Sesar et al. (2006) using recalibrated photometric POSS and SDSS measurements and by Morales-Rueda et al. (2006) using the Faint Sky Variability Survey data ( $\sim 1\%$ ).

Thanks to the multicolor nature of the SDSS photometry, and especially the  $u$ -band data, we can obtain robust classification of selected variable sources. The majority (2 out of 3) of variable sources are low-redshift ( $z < 2$ ) quasars, although they represent only 2% of all sources in the adopted  $g < 20.5$  flux-limited sample. We find that about 1 out of 4 of the variable stars are RR Lyrae stars, and that only 0.5% of the stars from the main stellar locus are variable at the 0.05 mag level.

The distribution of  $\gamma(g)$  for main stellar locus stars is bimodal, suggesting at least two, and perhaps more, different populations of variables. About one-third of the variable stars from the stellar locus show gray flux variations in the  $g$  and  $r$  bands [ $\sigma(g)/\sigma(r) \sim 1$ ] and positive light-curve skewness, suggesting variability caused by eclipsing systems. This population has an increased fraction of M-type stars.

RR Lyrae stars show the largest rms scatter in the  $u$  and  $g$  bands, followed by low-redshift quasars. The ratio of rms scatter in the  $g$  and  $r$  bands for RR Lyrae stars is  $\sim 1.4$ , in agreement with the Ivezić et al. (2000) results based on two-epoch photometry. The mean light-curve skewness for RR Lyrae stars is approximately  $-0.5$ , in agreement with Wils et al. (2006). We selected a sample of 634 candidate RR Lyrae stars with an estimated  $\geq 95\%$  completeness and  $\sim 70\%$  efficiency. Using these stars, we detected rich halo substructure out to distances of 100 kpc. The apparent “clumpiness” of the candidate RR Lyrae distribution increases with increasing radius, similar to CDM predictions by Bullock et al. (2001).

Low-redshift quasars show a dependence of  $\sigma(g)/\sigma(r)$  on redshift, consistent with discussions in Richards et al. (2002) and Wilhite et al. (2005). The light-curve skewness distribution for quasars is centered on zero in all photometric bands. We find that at least 90% of quasars are variable at the 0.03 mag level (rms) on timescales up to several years. This confirms that variability is as a good a method for finding low-redshift quasars at high ( $|b| > 30^\circ$ ) Galactic latitudes as UV excess color selection. The fraction of variable quasars at the  $\geq 0.1$  mag level obtained here (30%; see Fig. 14) is comparable to 36% found by Rengstorff et al. (2006).

The multiple photometric observations obtained by the SDSS represent an excellent data set for estimating the impact of surveys such as the LSST on studies of the variable sky. Our results indicate that the LSST will obtain well-sampled 2% accurate multicolor light curves for  $\sim 2$  million low-redshift quasars and will discover at least 50 million variable stars. The number of variable stars discovered by the LSST will be of the same order as the number of *all* stars detected by the SDSS. With about 1000 data points in six photometric bands, it will be possible to recognize

<sup>28</sup> Available at [http://www.lsst.org/Science/lst\\_baseline.shtml](http://www.lsst.org/Science/lst_baseline.shtml).

<sup>29</sup> The LSST Exposure Time Calculator is available at <http://www.lsst.org>.



and classify variable objects using light-curve moments of higher order than the skewness discussed here, including light-curve folding for periodic variables.

This work was supported by an NSF-sponsored research stipend (NVO 8601-06862) obtained from NVO through The Johns Hopkins University. We also acknowledge support by NSF grant AST 05-51161 to the LSST for design and development activity.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web site is <http://www.sdss.org>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, the Astrophysical Institute Potsdam, the University of Basel, the University of Cambridge, Case Western Reserve University, the University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences, Los Alamos National Laboratory, the Max Planck Institute for Astronomy, the Max Planck Institute for Astrophysics, New Mexico State University, Ohio State University, the University of Pittsburgh, the University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

#### REFERENCES

- Adelman-McCarthy, J. K., et al. 2007, *ApJS*, 172, in press
- Akerlof, C., et al. 2000, *AJ*, 119, 1901
- Alcock, C., et al. 2001, *ApJS*, 136, 439
- Becker, A. C., et al. 2004, *ApJ*, 611, 418
- Beck-Winchatz, B., & Anderson, S. F. 2007, *MNRAS*, 374, 1506
- Bell, E. F. et al. 2007, *ApJ*, submitted (arXiv:0706.0004)
- Bullock, J. S., Kravtsov, A. V., & Weinberg, D. H. 2001, *ApJ*, 548, 33
- de Vries, W. H., Becker, R. H., & White, R. L. 2003, *AJ*, 126, 1217
- de Vries, W. H., Becker, R. H., White, R. L., & Loomis, C. 2005, *AJ*, 129, 615
- Fan, X. 1999, *AJ*, 117, 2528
- Finlator, K., et al. 2000, *AJ*, 120, 2615
- Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, *AJ*, 111, 1748
- Groot, P. J., et al. 2003, *MNRAS*, 339, 427
- Gunn, J. E., et al. 1998, *AJ*, 116, 3040
- . 2006, *AJ*, 131, 2332
- Hawkins, M. R. S. 1983, *MNRAS*, 202, 571
- Hawkins, M. R. S., & Veron, P. 1995, *MNRAS*, 275, 1102
- Hawley, S. L., et al. 2002, *AJ*, 123, 3409
- Helmi, A., et al. 2003, *ApJ*, 586, 195
- Hoffmeister, C., Richter, G., & Wenzel, W. 1985, *Variable Stars* (New York: Springer)
- Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, *AJ*, 122, 2129
- Ivezić, Ž., Vivas, A. K., Lupton, R. H., & Zinn, R. 2005, *AJ*, 129, 1096
- Ivezić, Ž., et al. 2000, *AJ*, 120, 963
- . 2003, *Mem. Soc. Astron. Italiana*, 74, 978
- . 2004a, *Astron. Nachr.*, 325, 583
- . 2004b, in *IAU Symp.* 222, *The Interplay among Black Holes, Stars, and ISM in Galactic Nuclei*, ed. T. S. Bergmann, L. C. Ho, & H. R. Schmitt (Cambridge: Cambridge Univ. Press), 525
- . 2004c, in *ASP Conf. Ser.* 311, *AGN Physics with the Sloan Digital Sky Survey*, ed. G. T. Richards & P. B. Hall (San Francisco: ASP), 437
- . 2007, *AJ*, 134, 973
- Jurić, M., et al. 2007, *AJ*, submitted (astro-ph/0510520)
- Kaiser, N., et al. 2002, *Proc. SPIE*, 4836, 154
- Kawaguchi, T., Mineshige, S., Umemura, M., & Turner, E. L. 1998, *ApJ*, 504, 671
- Kholopov, P. N., et al. 1988, *General Catalog of Variable Stars* (4th ed.; Moscow: Nauka)
- Kinney, A. L., Bohlin, R. C., Blades, J. C., & York, D. G. 1991, *ApJS*, 75, 645
- Koo, D. C., Kron, R. G., & Cudworth, K. M. 1986, *PASP*, 98, 285
- Lenz, D. D., Newberg, J., Rosner, R., Richards, G. T., & Stoughton, C. 1998, *ApJS*, 119, 121
- Lupton, R. H., Ivezić, Ž., Gunn, J. E., Knapp, G. R., Strauss, M. A., & Yasuda, N. 2002, *Proc. SPIE*, 4836, 350
- Martini, P., & Schneider, D. P. 2003, *ApJ*, 597, L109
- Matthews, T. A., & Sandage, A. R. 1963, *ApJ*, 138, 30
- Monet, D. G., et al. 2003, *AJ*, 125, 984
- Morales-Rueda, L., Groot, P. J., Augusteijn, T., Nelemans, G., Vreeswijk, P. M., & van den Besselaar, E. J. M. 2006, *MNRAS*, 371, 1681
- National Research Council. 2001, *Astronomy and Astrophysics in the New Millennium* (Washington: Natl. Acad. Press)
- Pereyra, N. A., Vanden Berk, D. E., & Turnshek, D. A. 2006, *ApJ*, 642, 87
- Pier, J. R., Munn, J. A., Hindsley, R. B., Hennesy, G. S., Kent, S. M., Lupton, R. H., & Ivezić, Ž. 2003, *AJ*, 125, 1559
- Pojmański, G. 2002, *Acta Astron.*, 52, 397
- Rengstorf, A. W., Brunner, R. J., & Wilhite, B. C. 2006, *AJ*, 131, 1923
- Rengstorf, A. W., et al. 2004, *ApJ*, 606, 741
- Richards, G. T., et al. 2002, *AJ*, 123, 2945
- Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, *A&A*, 409, 523
- Schlegel, D., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schneider, D. P., et al. 2005, *AJ*, 130, 367
- Scranton, R., et al. 2002, *ApJ*, 579, 48
- Sesar, B., et al. 2006, *AJ*, 131, 2801
- Sirko, E., et al. 2004, *AJ*, 127, 899
- Smith, H. A. 1995, *RR Lyrae Stars* (Cambridge: Cambridge Univ. Press)
- Smith, J. A., et al. 2002, *AJ*, 123, 2121
- Smolčić, V., et al. 2004, *ApJ*, 615, L141
- Stoughton, C., et al. 2002, *AJ*, 123, 485
- Strateva, I., et al. 2001, *AJ*, 122, 1861
- Trevese, D., Kron, R. G., & Bunone, A. 2001, *ApJ*, 551, 103
- Tucker, D., et al. 2006, *Astron. Nachr.*, 327, 821
- Tyson, J. A. 2002, *Proc. SPIE*, 4836, 10
- Udalski, A., Żebruń, K., Szymanski, M., Kubiak, M., Soszyński, I., Szweczyk, O., Wyrzykowski, L., & Pietrzynski, G. 2002, *Acta Astron.*, 52, 115
- van den Bergh, S., Herbst, E., & Pritchett, C. 1973, *AJ*, 78, 375
- Vanden Berk, D. E., et al. 2004, *ApJ*, 601, 692
- Vivas, A. K., et al. 2004, *AJ*, 127, 1158
- Walker, A. R. 2003, *Mem. Soc. Astron. Italiana*, 74, 999
- Wilhite, B. C., Vanden Berk, D. E., Kron, R. G., Schneider, D. P., Pereyra, N., Brunner, R. J., Richards, G. T., & Brinkmann, J. V. 2005, *ApJ*, 633, 638
- Wils, P., Lloyd, C., & Bernhard, K. 2006, *MNRAS*, 368, 1757
- Woźniak, P. R., Udalski, A., Szymanski, M., Kubiak, M., Pietrzynski, G., Soszyński, I., & Żebruń, K. 2002, *Acta Astron.*, 52, 129
- Woźniak, P. R., et al. 2004, *AJ*, 127, 2436
- Yanny, B., et al. 2000, *ApJ*, 540, 825
- York, D. G., et al. 2000, *AJ*, 120, 1579
- Żebruń, K., et al. 2001, *Acta Astron.*, 51, 317